

Information Transmission using the Nonlinear Fourier Transform, Part I: Mathematical Tools

Mansoor I. Yousefi and Frank R. Kschischang *Fellow, IEEE*

Abstract—The nonlinear Fourier transform (NFT), a powerful tool in soliton theory and exactly solvable models, is a method for solving integrable partial differential equations governing wave propagation in certain nonlinear media. The NFT decorrelates signal degrees-of-freedom in such models, in much the same way that the Fourier transform does for linear systems. In this paper, this observation is exploited for data transmission over integrable channels such as optical fibers, where pulse propagation is governed by the nonlinear Schrödinger equation. In this transmission scheme, which can be viewed as a nonlinear analogue of orthogonal frequency division multiplexing commonly used in linear channels, information is encoded in the spectral amplitudes associated with nonlinear frequencies. Unlike most other fiber-optic transmission schemes, this technique deals with both dispersion and nonlinearity directly and unconditionally without the need for dispersion or nonlinearity compensation methods. This paper explains the mathematical tools that underlie the method.

Index Terms—Fiber-optic communications, nonlinear Fourier transform, integrable channels, Lax pairs, Zakharov-Shabat spectral problem.

I. INTRODUCTION

AN EVOLUTION equation is a partial differential equation (PDE) for an unknown function $q(t, z)$ of the form

$$q_z = K(q), \quad (1)$$

where $K(q)$ is an expression involving q and its derivatives with respect to t . In (1) and throughout this paper, subscripts are used to denote partial derivatives with respect to the corresponding variable.

The evolution equation (1) is a PDE in $1 + 1$ dimensions, i.e., one variable $t \in \mathbb{R}$ represents a temporal dimension and one variable $z \geq 0$ represents a spatial dimension. In most mathematical literature, the roles of z and t are interchanged, so rather than a spatial evolution as in (1), mathematicians study temporal evolution.

An important example of an evolution equation, and the one that motivates this work, is the stochastic nonlinear Schrödinger (NLS) equation governing the complex envelope of a narrowband signal in an optical fiber in the presence of

noise, given by [1], [2]

$$\frac{\partial Q(\tau, \ell)}{\partial \ell} + \frac{j\beta_2}{2} \frac{\partial^2 Q(\tau, \ell)}{\partial \tau^2} = j\gamma Q(\tau, \ell)|Q(\tau, \ell)|^2 + V(\tau, \ell), \quad 0 \leq \ell \leq \mathcal{L}. \quad (2)$$

Here ℓ denotes distance (in km) along the fiber; the transmitter is located at $\ell = 0$, and the receiver is located at $\ell = \mathcal{L}$. The symbol τ represents retarded time, measured in seconds, i.e., $\tau = t - \beta_1 \ell$ where t is ordinary time and β_1 is a constant, and $Q(\tau, \ell)$ is the complex envelope of the signal propagating in the fiber. The coefficient β_2 , measured in s^2/km , is called the chromatic dispersion coefficient, while γ , measured in $\text{W}^{-1}\text{km}^{-1}$, is the nonlinearity parameter. Finally, $V(\tau, \ell)$ is bandlimited white Gaussian noise with in-band spectral density σ_0^2 ($\text{W}/(\text{km} \cdot \text{Hz})$) and autocorrelation

$$E\{V(\tau, \ell)V^*(\tau', \ell')\} = \sigma_0^2 \delta_W(\tau - \tau')\delta(\ell - \ell'),$$

where $\delta_W(x) = 2W\text{sinc}(2Wx)$. It is assumed that $Q(\tau, \ell)$ is bandlimited to W for all ℓ , $0 \leq \ell \leq \mathcal{L}$.

The stochastic NLS equation (2) models both chromatic dispersion (captured by the $j\beta_2 \partial^2 Q / \partial \tau^2$ term), which is responsible for temporal broadening, and the Kerr nonlinearity (captured by the $\gamma|Q|^2 Q$ term), which is responsible for spectral broadening. Pulse propagation is governed by the tension between these effects and can be linearly dominated, nonlinearly dominated, or solitonic (in which case the two effects are balanced). The NLS equation defines a nonlinear dispersive waveform channel from $Q(\tau, 0)$ at the transmitter to $Q(\tau, \mathcal{L})$ at the receiver.

Such a nonlinear dispersive waveform channel is a major departure from the classical additive white Gaussian noise and wireless fading channels in terms of analytical difficulty. Here the signal degrees-of-freedom couple together via the nonlinearity and dispersion in a complicated manner, making it difficult to establish the channel input-output map, even deterministically. Most current approaches assume a linearly-dominated regime of operation, or consider the nonlinearity as a small perturbation, and are geared towards managing and suppressing the (detrimental) effects of the nonlinear and dispersive terms. Inline dispersion management, digital backpropagation, and other forms of electronic pre- and post-compensation belong to this class of methods (see [3]–[10] and references therein).

In this paper we adopt a different philosophy. Rather than treating nonlinearity and dispersion as nuisances, we

Submitted for publication on February 16, 2012. The authors are with the Edward S. Rogers Sr. Dept. of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 1A4, Canada. Email: {mansoor, frank}@comm.utoronto.ca.

seek a transmission scheme that is fundamentally compatible with these effects. We effectively “diagonalize” the nonlinear Schrödinger channel with the help of the nonlinear Fourier transform (NFT), a powerful tool for solving *integrable* nonlinear dispersive partial differential equations (PDEs) [11], [12]. The NFT uncovers linear structure hidden in the one-dimensional cubic nonlinear Schrödinger equation, and can be viewed as a generalization of the (ordinary) Fourier transform to certain nonlinear systems.

With the help of the nonlinear Fourier transform, we are able to represent a signal by its discrete and continuous nonlinear spectra. While the signal propagates along the fiber based on the complicated NLS equation, the action of the channel on its spectral components is given by simple independent linear equations. Just as the (ordinary) Fourier transform converts a linear convolutional channel $y(t) = x(t) * h(t)$ into a number of parallel scalar channels, the nonlinear Fourier transform converts a nonlinear dispersive channel described by a *Lax convolution* (see Sec. III) into a number of parallel scalar channels. This suggests that information can be encoded (in analogy with orthogonal frequency-division multiplexing) in the nonlinear spectra.

The nonlinear Fourier transform is intertwined with the existence of soliton solutions to the NLS equation. Solitons are pulses that retain their shape (or return periodically to their initial shape) during propagation, and can be viewed as system eigenfunctions, similar to the complex exponentials $e^{j\omega t}$, which are eigenfunctions of linear systems. An arbitrary waveform can be viewed as a combination of solitons, associated with the discrete nonlinear spectrum, and a non-solitonic (radiation) component, associated with the continuous nonlinear spectrum.

The goal of this first article is to introduce the mathematical tools that underlie this approach to information transmission. These tools are sufficiently general to encompass not only the nonlinear Schrödinger equation, but also other completely integrable nonlinear dispersive PDEs. Thus, the transmission scheme described here can also be applied to any channel model in this general class. These tools are also described in mathematics and physics (see, e.g., [11]–[13]); here we attempt to extract those aspects of the theory that are relevant to the engineering aspects of the information transmission problem. In subsequent articles, we will describe useful numerical methods and give performance results for actual fiber-optic systems.

We will find it convenient to work with the nonlinear Schrödinger equation (2) in a normalized form. By changing variables

$$q = \frac{Q}{\sqrt{P}}, \quad z = \frac{l}{\mathcal{L}}, \quad t = \frac{\tau}{T_0}, \quad (3)$$

with $T_0 = \sqrt{\frac{|\beta_2|\mathcal{L}}{2}}$ and $P = \frac{2}{\gamma\mathcal{L}}$, we get the normalized NLS equation

$$jq_z(t, z) = q_{tt} + 2|q(t, z)|^2 q(t, z) + v(t, z), \quad (4)$$

where

$$\mathbb{E} \{v(t, z)v^*(t', z')\} = \frac{\sigma_0^2}{PT_0} \delta_W(t - t') \delta(z - z').$$

Throughout this paper, (4) will be our primary and motivating illustrative example.

II. A BRIEF HISTORY OF THE NONLINEAR FOURIER TRANSFORM

The nonlinear Fourier transform (also known as the inverse scattering transform or IST) is a method for solving certain nonlinear dispersive partial differential equations (PDEs). These are integrable PDEs, *i.e.*, nonlinear differential equations exhibiting certain hidden linearity. There are several integrable equations having physical significance, among which is the NLS equation. The IST method was a result of extensive efforts in theoretical physics and applied mathematics in 1960s, closely associated with the notion of solitons and is often used to predict their existence and properties in integrable models [11], [12].

In the 1950s, in one of the first dynamical-systems simulations performed on a computer [14], Fermi, Pasta and Ulam performed a numerical experiment to understand why solids have finite heat conductivity. They modeled the solid as a lattice with point masses at the lattice points coupled with springs each having a quadratic nonlinearity. To their surprise, rather than observing an equipartition of energy among all modes, energy cycled periodically among a few low-order modes. Such behavior implies that the nonlinear oscillator behaves somehow linearly.

In the 1960s, Zabusky and Kruskal showed that equation of motion for the Fermi-Pasta-Ulam lattice in the continuum limit is a remarkable PDE called the Korteweg-de Vries (KdV) equation [15], known in the study of water waves. The KdV equation for the evolution of a real-valued pulse $q(t, z)$ as a function of time t and distance z is

$$q_z = qq_t + q_{ttt}. \quad (5)$$

Zabusky and Kruskal found that (5) has pulse-like solutions whose shape is preserved (or varies periodically) during propagation. Furthermore, they made the surprising observation that when two such pulses are launched towards each other, despite their nonlinear interaction, they pass through each other without changing their shape. Zabusky and Kruskal coined the term *soliton* for such solutions, in recognition of their particle-like properties [15].

The spectacular properties of these solutions greatly excited the mathematics and physics communities and many researchers started to study solitons. In a celebrated paper [16], Gardner, Greene, Kruskal and Miura uncovered some of the deep structure underlying the KdV equation which is responsible for solitons and their unusual properties. The authors of [16] were studying the celebrated linear Schrödinger equation from quantum mechanics, given by

$$\psi_z(t, z) = \psi_{tt}(t, z) + g(t, z)\psi(t, z)/3, \quad (6)$$

where $\psi(t, z)$ is the wave-function and $g(t, z)$ is an external potential. They found that if one takes the solution $q(t, z)$ of the KdV equation (5) as the external potential in (6), then the eigenvalues of the Schrödinger operator

$$H = \frac{\partial^2}{\partial t^2} + g(t, z)/3 \quad (7)$$

remain invariant during the evolution in z ! Based on this critical observation, they developed a method to recover the external potential $g(t, z) = q(t, z)$ by solving an inverse problem for (6). The method analytically predicts soliton solutions for the KdV equation, as observed earlier by Zakusky and Kruskal through numerical computations. They had in fact found the IST for the special case of the KdV equation.

It was not immediately clear if the method developed in [16] could be generalized to other nonlinear PDEs, since it is not obvious if there exists a certain auxiliary operator, like the Schrödinger operator H , whose eigenvalues are preserved during the evolution. In a landmark paper published in 1968 [17], Lax put the theory on a firm mathematical footing. In particular, he established the mathematical relationship between the auxiliary operators with invariant eigenvalues (now called Lax pairs) and the original nonlinear equation. Once a Lax pair for a nonlinear PDE is found, a method along the lines of [16] can be applied to solve that PDE.

Shortly afterwards, in 1972, Zakharov and Shabat found a Lax pair for the NLS equation in one spatial dimension [1], and thus established that this equation, too, could be solved in the same manner. Details of this method for the NLS equation were subsequently developed by Ablowitz and others (see [11] and references therein), who also referred to this scheme as the “nonlinear Fourier transform”. After these discoveries from the 1960s and 70s, research into solitons became an established area of research, lying at the intersection of applied mathematics and nonlinear physics.

Nonlinear PDEs solvable by the NFT are called integrable equations or exactly solvable models [18]. These are usually Hamiltonian systems having an infinite number of conserved quantities, and include the KdV, NLS, modified KdV, and sine-Gordon equations, the Toda lattice, among others [11], [12]. These equations all exhibit similar properties, including the existence of soliton solutions. Signal processing problems (e.g., detection and estimation) involving soliton signals in the Toda lattice have been considered by Singer [19].

III. CANONICAL LAX FORM FOR EXACTLY SOLVABLE MODELS

A. Lax Pairs and Evolution Equations

We wish to consider linear differential operators whose eigenvalues are invariant during an evolution [17]. More precisely, we consider such operators defined in terms of a signal $q(t, z)$ where the eigenvalues of the operator remain constant even as q evolves (in z) according to some evolution equation.

To facilitate the discussion, it is useful to imagine a linear operator represented as a matrix; however, we must keep in mind that, when moving from finite-dimensional spaces to infinite-dimensional spaces (of e.g., functions and operators), some results do not carry over necessarily. The relevant properties of linear operators needed for this paper are reviewed in Appendix A.

Let $L(z)$ be a square matrix whose entries are functions of z . Clearly, the eigenvalues of this matrix are in general functions of z too. However, for some matrices, it might be

the case that while the entries of the matrix change with z , the eigenvalues remain constant (independent of z). Such a matrix, if diagonalizable, should be similar to a constant diagonal matrix Λ , i.e., $L(z) = G(z)\Lambda G^{-1}(z)$, for some similarity transformation $G(z)$.

This idea generalizes to operators. Let \mathcal{H} be a Hilbert space, let \mathcal{D} be some domain that is dense in \mathcal{H} , and let $L(z) : \mathcal{D} \rightarrow \mathcal{H}$ be a family of bounded linear operators indexed by a parameter z [20]. If the eigenvalues of $L(z)$ do not depend on z , then we refer to $L(z)$ as an *isospectral* family of operators. If diagonalizable, it follows that for each z , $L(z)$ is similar to a multiplication operator Λ (the operator equivalent of a diagonal matrix; see Appendix A), i.e., $L(z) = G(z)\Lambda G^{-1}(z)$, for some operator $G(z)$.

Assuming that $L(z)$ varies smoothly with z , we can consider the rate of the change (with respect to z) of $L(z)$. We have

$$\begin{aligned} \frac{dL(z)}{dz} &= G' \Lambda G^{-1} + G \Lambda (-G^{-1} G' G^{-1}) \\ &= G' G^{-1} (G \Lambda G^{-1}) - (G \Lambda G^{-1}) G' G^{-1} \\ &= M(z) L(z) - L(z) M(z) = [M, L], \end{aligned} \quad (8)$$

where $G' = dG(z)/dz$, $M = G' G^{-1}$, and $[M, L] \triangleq ML - LM$ is the *commutator bracket*. In other words, every diagonalizable isospectral operator $L(z)$ satisfies the differential equation (8).

Conversely, suppose $M(z)$ is given and the (unknown) diagonalizable operator $L(z)$ evolves according to (8) with initial condition $L(0) = G_0 \Lambda_0 G_0^{-1}$. Let $G(z)$ be the (unique invertible) solution to $G' = MG$ with $G(0) = G_0$. One can easily verify that $L(z) = G(z) \Lambda_0 G(z)^{-1}$ satisfies (8). Since the solution to a first-order differential equation is unique [21], we see that $L(z)$ is an isospectral family.

The characterization of isospectral operators is therefore summarized in the following lemma [17].

Lemma 1. *Let $L(z)$ be a diagonalizable operator. Then $L(z)$ is isospectral if and only if it satisfies*

$$\frac{dL}{dz} = [M, L], \quad (9)$$

for some operator M . If L is self-adjoint (so that L is unitarily equivalent to a multiplication operator; i.e., $L = G \Lambda G^*$), then M must be skew-Hermitian, i.e., $M^* = -M$.

Proof: The proof was outlined above. The skew-Hermitian property of M can be shown by differentiating $GG^* = I$. ■

It is important to note that L and M do not have to be independent and can depend on a common parameter, e.g., a function $q(t, z)$, as illustrated in Fig. 1. The isospectral property of the solution is unchanged. The commutator bracket $[M, L]$ in (9) can create nonlinear evolution equations for $q(t, z)$ in the form

$$\frac{\partial q}{\partial z} = K(q),$$

where $K(q)$ is some, in general nonlinear, function of $q(t, z)$ and its time derivatives. An example of this is the KdV equation.

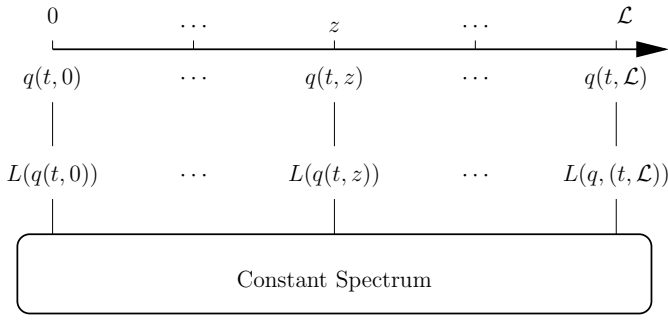


Fig. 1. An isospectral flow: the spectrum of L is held invariant even as $q(t, z)$ evolves.

Example 1 (KdV equation). Let $D = \frac{\partial}{\partial t}$ denote the time-derivative operator, and let $q(t, z)$ be a real-valued function. Finally, let

$$L = D^2 + \frac{1}{3}q, \text{ and let } M = 4 \left(D^3 + \frac{1}{4}Dq + \frac{1}{4}qD \right). \quad (10)$$

The Lax equation $L_z = [M, L]$ is easily simplified to

$$\frac{1}{3}q_z - \frac{1}{3}(q_{ttt} + qq_t) + (\text{some terms})D \equiv \mathbf{0}, \quad (11)$$

where $\mathbf{0}$ is the zero operator. The zero-order term of this equation as a polynomial in D , which must be zero, produces the KdV equation $q_z = q_{ttt} + qq_t$. Using $(D^k)^* = (-1)^k D$ (provable by integration by parts), it is easy to see that L and M are self-adjoint and skew-Hermitian, respectively, since they are real valued and involve even and odd powers of D , respectively. The eigenvalues of L are thus preserved from Lemma 1. Note that the L operator in this example is precisely the (linear) Schrödinger operator H given in (7). \square

Definition 1. A pair of operators L and M , depending on z , are called a Lax pair (L, M) if they satisfy (9). Following Lemma 1, the eigenvalues of the L operator are independent of z .

B. The Zero-Curvature Condition

The eigenvalues of the operator L , which are constant in an isospectral flow, are defined via

$$Lv = \lambda v. \quad (12)$$

Taking the z derivative of (12) and using the Lax equation $L_z = [M, L]$, we obtain $(L - \lambda I)(v_z - Mv) = 0$. Since $L - \lambda I$ vanishes only on eigenvectors of L , it must be that $v_z - Mv$ is proportional to an eigenvector, i.e., $v_z - Mv = \alpha v$, $\alpha \in \mathbb{C}$. The choice of α does not influence the results of future sections; thus for simplicity we set $\alpha = 0$. It follows that an eigenfunction $v(t, z)$ evolves based on the linear equation

$$v_z = Mv. \quad (13)$$

Furthermore, it is often more convenient to re-write (12) as

$$v_t = Pv, \quad (14)$$

for some operator P . The relationship between P and L can be derived (if necessary) by combining $(DI - P)v = 0$ with $(L - \lambda I)v = 0$, obtaining

$$P = \Sigma(L - \lambda I) + DI, \quad (15)$$

where Σ is some invertible operator, and $D = \frac{\partial}{\partial t}$ as in Example 1.

Combining equations (13) and (14) by using the equality of mixed derivatives, i.e., $v_{tz} = v_{zt}$, the Lax equation (9) is reduced to the *zero-curvature condition* [12]

$$P_z - M_t + [P, M] = 0. \quad (16)$$

Note that the nonlinear equation derived from (16) results as a compatibility condition between the two linear equations (13) and (14). This shows that certain nonlinear equations possess a “hidden linearity” in the form of (13) and (14).

Following the work of Zakharov and Shabat on the NLS equation [1], Ablowitz et al. [22] suggested that for many equations of practical significance, the operator P can be fixed as

$$P = \begin{pmatrix} -j\lambda & r(t, z) \\ s(t, z) & j\lambda \end{pmatrix}, \quad (17)$$

where $r(t, z)$ and $s(t, z)$ are functions—depending on $q(t, z)$ —to be determined to produce a given nonlinear evolution equation. From (15) this corresponds to the L operator

$$L = j \begin{pmatrix} D & -r(t, z) \\ s(t, z) & -D \end{pmatrix},$$

with $\Sigma = \text{diag}(j, -j)$. In this case, both L and P operate on 2×1 vector functions.

Equation (14), with P as in (17), is known as the *AKNS system* (after the authors of [22]) and is central in the study of the nonlinear Fourier transform [12]. The important special case where $r(t, z) = q(t, z)$ and $s(t, z) = -q^*(t, z)$ is generally known as the *Zakharov-Shabat system*. We will refer to (12), (13) and (14) as the L -, M -, and P -equations, respectively. Throughout this paper, we will assume that the domain of L , P , and M are subsets of the Hilbert space $L^2(\mathbb{R})$, denoted here by \mathcal{H} , depending on the particular structure of that operator.

Example 2 (sine-Gordon equations). Let $r(t, z) = -s(t, z) = \frac{1}{2}q_t(t, z)$ in (17) and let

$$M = \frac{j}{4\lambda} \begin{pmatrix} \cos(q) & -\sin(q) \\ -\sin(q) & -\cos(q) \end{pmatrix}. \quad (18)$$

Then the zero-curvature equation is simplified to $q_{tz} = \sin(q)$. Taking $r = s = \frac{1}{2}q_t$ and M as

$$M = \frac{j}{4\lambda} \begin{pmatrix} \cosh(q) & -\sinh(q) \\ \sinh(q) & -\cosh(q) \end{pmatrix}. \quad (19)$$

gives $q_{tz} = \sinh(q)$. \square

Example 3 (Nonlinear Schrödinger equation). Take $r = q$, $s = -q^*$ and

$$M = \begin{pmatrix} 2j\lambda^2 - j|q(t, z)|^2 & -2\lambda q(t, z) - jq_t(t, z) \\ 2\lambda q^*(t, z) - jq_t^*(t, z) & -2j\lambda^2 + j|q(t, z)|^2 \end{pmatrix}.$$

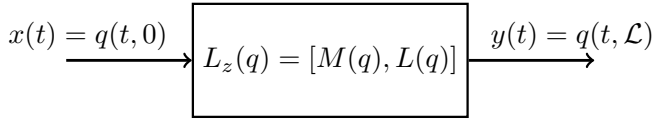


Fig. 2. A system defined by Lax convolution.

The zero-curvature equation is simplified to $j q_z(t, z) = q_{tt}(t, z) + 2|q(t, z)|^2 q(t, z)$. \square

Example 4 (KdV equation revisited). Let $r = \frac{2}{\sqrt{3}}q$, $s = -\frac{1}{4\sqrt{3}}$ and

$$M = \begin{pmatrix} 4j\lambda^3 - \frac{j\lambda q}{3} + \frac{q_t}{6} & -\frac{8\lambda^2 q}{\sqrt{3}} - \frac{4}{\sqrt{3}}j\lambda q_t + \frac{2q^2}{3\sqrt{3}} + \frac{2q_{tt}}{\sqrt{3}} \\ \frac{\lambda^2}{\sqrt{3}} - \frac{q}{12\sqrt{3}} & -4j\lambda^3 + \frac{j\lambda q}{3} - \frac{q_t}{6} \end{pmatrix}. \quad (20)$$

The zero-curvature equation leads to $q_z = qq_t + q_{tt}$. \square

It should be noted from Examples 1 and 4 that the choice of L and M giving rise to a given nonlinear equation is not unique. Obviously one can scale L , or add a constant to L or M . In addition, both the Lax equation (9) and (12) are unchanged under orthogonal transformations, i.e., replacing L and M with $\Sigma L \Sigma^T$ and $\Sigma M \Sigma^T$, respectively, where Σ is a (constant) orthogonal matrix, i.e., $\Sigma^T \Sigma = I$. Note further that it may be possible to choose two Lax pairs (L_1, M_1) and (L_2, M_2) for a given equation such that L_1 is self-adjoint and L_2 is not self-adjoint. The eigenvalues of L_1 and L_2 are, respectively, real and complex; see Appendix A.

C. Lax Convolution and Integrable Communication Channels

Linear systems traditionally have been described by linear constant coefficient differential equations. An example is the one-dimensional heat equation $q_z = c^2 q_{tt}$, where c is the diffusion coefficient and $q(t, z)$ represents the heat profile across a rod extending in *space* t , as *time* z goes on. From a systems point of view, this defines a linear time-invariant system from input $x(t) = q(t, 0)$ at $z = 0$ to the output $y(t) = q(t, \mathcal{L})$ at some $z = \mathcal{L}$. The role of z is therefore just a parameter and once fixed (to $z = \mathcal{L}$), the system is described by an impulse response (or a Green function) $h(t; \mathcal{L})$, representing the underlying (linear) convolution.

Following this analogy, we wish to define a system in terms of a Lax pair (L, M) . Here, L and M are parametrized by a waveform $q(t, z)$. Such a system accepts a waveform $x(t) = q(t, 0)$ at its input and produces a waveform $y(t) = q(t, \mathcal{L})$ at its output, according to the evolution equation induced by $L_z = [M, L]$. The time-domain input-output map is thus given by an evolution equation of the form $q_z = K(q)$, obtainable from the Lax equation (9) (or its equivalent (16)). We refer to such a system as an *integrable system*. Note that an integrable system is completely characterized by the two operators (L, M) and the parameter $z = \mathcal{L}$, independent of the signals. We denote such a system using the triple $(L, M; \mathcal{L})$.

Definition 2 (Lax Convolution). We refer to the action of an integrable system $S = (L, M; \mathcal{L})$ on the input $q(t, 0)$ as the Lax convolution of q with S . We write the system output as $q(t, \mathcal{L}) = q(t, 0) * (L, M; \mathcal{L})$. See Fig. 2.

Definition 3 (Integrable communication channels). A waveform communication channel $C : x(t) \times v(t, z) \rightarrow y(t)$ with inputs $x(t) \in L^1(\mathbb{R})$ and space-time noise $v(t, z) \in L^2(\mathbb{R}, \mathbb{R}^+)$, and output $y(t) \in L^1(\mathbb{R})$, is said to be *integrable* if the noise-free channel is an integrable system.

Note that noise can be introduced into an integrable system in a variety of ways. In what follows, we assume that the signal-to-noise ratio is not too small, so that the stochastic system may justifiably be considered as a perturbation of the deterministic system. Furthermore, in this paper we limit ourselves to integrable channels with additive noise, i.e.,

$$q_z = K(q) + v(t, z), \quad q(t, 0) = x(t), \quad q(t, \mathcal{L}) = y(t),$$

where $v(t, z)$ is distributed band-limited noise. Note that this model is *not* in general equivalent to one in which noise is added (in lumped fashion) at the channel output. Here, the noise is distributed in space, and so interacts with the signal (in a potentially very complicated manner) according to the given evolution equation.

In this paper, we develop a scheme for communication over integrable channels. By various choices of Lax pair (L, M) , one can construct a variety of interesting channel models, mostly nonlinear, which go beyond the linear channel models typically studied in data communications. Interestingly, some linear channels can also be analyzed using this nonlinear spectral approach advocated here [23]. The central application (and motivation) for this work is fiber-optic communication, in which the channel model is given by the nonlinear Schrödinger equation (2), and for which a Lax pair was given in Example 3. In what follows, we first discuss the deterministic (noise-free) case and later treat noise as a perturbation of the deterministic system.

IV. NONLINEAR FOURIER TRANSFORM

In this section, we assume that a function $q(t, \cdot) \in L^1(\mathbb{R})$ is given, and we define its nonlinear Fourier transform with respect to a given Lax operator L . As the notation $q(t, \cdot)$ implies, in this and the next section, the variable z can take on any value in the range $[0, \mathcal{L}]$ and is irrelevant in the forward and inverse transforms. We shall therefore omit the index z in what follows.

As previously noted, for concreteness we carry through the development of the NFT for the case of the NLS equation (4), for which the P - and M -equations are given by

$$v_t = P(\lambda, q)v = \begin{pmatrix} -j\lambda & q(t, z) \\ -q^*(t, z) & j\lambda \end{pmatrix} v, \quad (21)$$

$$v_z = M(\lambda, q)v = \begin{pmatrix} 2j\lambda^2 - j|q(t, z)|^2 & -2\lambda q(t, z) - jq_t(t, z) \\ 2\lambda q^*(t, z) - jq_t^*(t, z) & -2j\lambda^2 + j|q(t, z)|^2 \end{pmatrix} v, \quad (22)$$

respectively. Here the P -equation is the Zakharov-Shabat system.

The nonlinear Fourier transform (and its inverse) are defined solely in terms of the temporal P -equation (21). The spatial M -equation (22) is used in a brief, and critical, propagation step, but plays no role in the definition of the nonlinear Fourier transform.

If for a given $\lambda \in \mathbb{C}$, the operator $L - \lambda I$, here equivalent to $DI - P$ representing (21), is not invertible, we say that λ belongs to the spectrum of L and $v(t, \lambda)$ represents its associated eigenfunction. In finite-dimensional Hilbert spaces of matrices the spectrum is a discrete finite set, called eigenvalues. This may no longer be true in infinite-dimensional spaces of operators, where the eigenvalues (if they exist) may only be one part of the spectrum. See Appendix A.

The nonlinear Fourier transform of a signal $x(t)$ with respect to an operator L in a Lax pair is defined via the spectral analysis of the L operator, which we consider next.

A. Canonical Eigenvectors and Spectral Coefficients

We wish to study solutions of (21), in which vectors $v(t)$ are considered as elements of the vector space \mathcal{H} . We begin by equipping the vector space \mathcal{H} with a symplectic bilinear form $\mathcal{H} \times \mathcal{H} \mapsto \mathbb{C}$, which, for any fixed value of $t \in \mathbb{R}$, is defined as

$$\langle v(t), w(t) \rangle_s = v_1(t)w_2(t) - v_2(t)w_1(t). \quad (23)$$

Let us also define the adjoint of any vector v in H as

$$\tilde{v}(t) = \begin{pmatrix} v_2^*(t) \\ -v_1^*(t) \end{pmatrix}.$$

The following properties hold true for all v and w in H :

- $\tilde{\tilde{v}} = -v$;
- $\langle v, v \rangle_s = 0$;
- $\langle v, w \rangle_s = -\langle w, v \rangle_s$;
- $\langle \tilde{v}, v \rangle_s = -\langle v, \tilde{v} \rangle_s = |v_1|^2 + |v_2|^2$;
- $\frac{d}{dt} \langle v, w \rangle_s = \langle v_t, w \rangle_s + \langle v, w_t \rangle_s$;
- for every 2×2 matrix A , $\langle Av, w \rangle_s + \langle v, Aw \rangle_s = \text{tr}(A) \langle v, w \rangle_s$.

There are generally infinitely many solutions v of (21) for a given $\lambda \in \mathbb{C}$, parametrized by the set of all possible boundary conditions. These solutions form a subspace E_λ of continuously differentiable 2×1 vector functions (an eigenspace).

Lemma 2. For all vectors $v(t)$ and $w(t)$ in E_λ ,

- 1) $\tilde{v} \in E_{\lambda^*}$, i.e., $\tilde{v}_t = P(\lambda^*, q)\tilde{v}$;
- 2) $\langle v(t), w(t) \rangle_s$ is a constant, independent of t ;
- 3) If $\langle v(t), w(t) \rangle_s \neq 0$, then v and w are linearly independent and form a basis for E_λ ;
- 4) $\dim(E_\lambda) = 2$.

Proof: Property 1) follows directly from (21). To see 2), note that $\frac{d}{dt} \langle v, w \rangle_s = \langle v_t, w \rangle_s + \langle v, w_t \rangle_s = \langle Pv, w \rangle_s + \langle v, Pw \rangle_s = \text{tr}(P) \langle v, w \rangle_s = 0$. To see 3), fix t and let $u(t) \in E_\lambda$, then $u(t) = a(t)v(t) + b(t)w(t)$ for some $a(t)$ and $b(t)$. Taking the symplectic inner product of both sides with w and v , we get $a(t) = \langle u, w \rangle_s / \langle v, w \rangle_s$ and $b(t) = \langle u, v \rangle_s / \langle w, v \rangle_s$. From Property 2, $\langle u, w \rangle_s$, $\langle v, w \rangle_s$, $\langle u, v \rangle_s$, and $\langle w, v \rangle_s$ are all independent of t . It follows that a and b are also independent of t . Finally, 4) follows from 3). \square

An important conclusion of Lemma 2 is that any two linearly independent solutions u and w of (21) provide a basis for the solution space. To choose two such solutions,

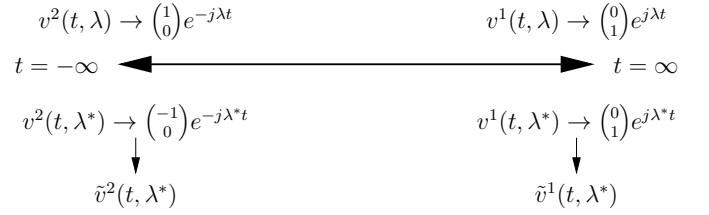


Fig. 3. Boundary conditions for the canonical eigenvectors.

we examine the behavior of the equation at large values of $|t|$. If we assume that $q(t, \cdot) \rightarrow 0$ as $|t| \rightarrow \infty$, then, as $|t| \rightarrow \infty$ (21) is reduced to

$$v_t \rightarrow \begin{pmatrix} -j\lambda & 0 \\ 0 & j\lambda \end{pmatrix} v, \quad \text{for large } |t|,$$

which has a general solution

$$v(t, \lambda) \rightarrow [\alpha e^{-j\lambda t}, \beta e^{j\lambda t}]^T, \quad \alpha, \beta \in \mathbb{C}.$$

Two possible boundary conditions, bounded in the upper half complex plane ($\Im(\lambda) > 0$), are

$$v^1(t, \lambda) \rightarrow \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{j\lambda t}, \quad t \rightarrow +\infty \quad (24)$$

$$v^2(t, \lambda) \rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-j\lambda t}, \quad t \rightarrow -\infty. \quad (25)$$

We solve (21) for a given λ under the boundary conditions (24)-(25), and denote the resulting solutions for all $t \in \mathbb{R}$ as $v^1(t, \lambda)$ and $v^2(t, \lambda)$. We can also solve (21) for λ^* under the adjoint boundary conditions

$$v^1(t, \lambda^*) \rightarrow \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{j\lambda^* t}, \quad t \rightarrow +\infty, \quad (26)$$

$$v^2(t, \lambda^*) \rightarrow \begin{pmatrix} -1 \\ 0 \end{pmatrix} e^{-j\lambda^* t}, \quad t \rightarrow -\infty,$$

giving rise to two solutions $v^1(t, \lambda^*)$ and $v^2(t, \lambda^*)$. From Lemma 2 we have that $\tilde{v}^1(t, \lambda^*)$ and $\tilde{v}^2(t, \lambda^*)$ are elements of E_λ . These four eigenfunctions $v^1(t, \lambda)$, $v^2(t, \lambda)$, $\tilde{v}^1(t, \lambda^*)$, and $\tilde{v}^2(t, \lambda^*)$, all of them elements of E_λ , are called *canonical eigenvectors*. Fig. 3 illustrates the canonical eigenvectors at their boundaries.

Lemma 3. Canonical eigenvectors satisfy:

- 1) $\langle \tilde{v}^1(t, \lambda^*), v^1(t, \lambda) \rangle_s = \langle \tilde{v}^2(t, \lambda^*), v^2(t, \lambda) \rangle_s = 1$;
- 2) $\{v^1(t, \lambda), \tilde{v}^1(t, \lambda^*)\}$ and $\{v^2(t, \lambda), \tilde{v}^2(t, \lambda^*)\}$ are independent sets in E_λ .

Proof: 1) Since $\langle \tilde{v}^1, v^1 \rangle_s$ is independent of t , using (24) and (26), $\langle \tilde{v}^1(t, \lambda^*), v^1(t, \lambda) \rangle_s = \langle \tilde{v}^1(+\infty, \lambda^*), v^1(+\infty, \lambda) \rangle_s = 1$. 2) Follows from 1) and Lemma 2. \square

Choosing $\tilde{v}^1(t, \lambda^*)$ and $v^1(t, \lambda)$ as a basis of E_λ , one can project $v^2(t, \lambda)$, $\tilde{v}^2(t, \lambda^*) \in E_\lambda$ on this basis to obtain

$$v^2(t, \lambda) = a(\lambda) \tilde{v}^1(t, \lambda^*) + b(\lambda) v^1(t, \lambda), \quad (27)$$

$$\tilde{v}^2(t, \lambda^*) = b^*(\lambda^*) \tilde{v}^1(t, \lambda^*) - a^*(\lambda^*) v^1(t, \lambda), \quad (28)$$

where $a(\lambda) = \langle v^2, v^1 \rangle_s$ and $b(\lambda) = \langle \tilde{v}^1, v^2 \rangle_s$. A crucial property, following from Lemma 2, is that $a(\lambda)$ and $b(\lambda)$ are time-independent. The time-independent complex scalars $a(\lambda)$

and $b(\lambda)$ are called the *nonlinear Fourier coefficients* [11], [12].

Since the nonlinear Fourier coefficients are time independent, to facilitate computing them, for simplicity we can send t to, e.g., $+\infty$ where $v^1(+\infty, \lambda)$ and $\tilde{v}^1(+\infty, \lambda^*)$ are known. The other two canonical eigenvectors v^2 and \tilde{v}^2 are then propagated from their boundary values $v^2(-\infty, \lambda)$ and $\tilde{v}^2(-\infty, \lambda^*)$ at $t = -\infty$ according to $v_t^2 = P(\lambda, q)v^2$ and $\tilde{v}_t^2 = P(\lambda^*, q)\tilde{v}^2$ to obtain $v^2(+\infty, \lambda)$ and $\tilde{v}^2(+\infty, \lambda^*)$. At this stage, we have available all four canonical eigenvectors at one time, namely,

$$\{v^2(+\infty, \lambda), \tilde{v}^2(+\infty, \lambda), v^1(+\infty, \lambda), \tilde{v}^1(+\infty, \lambda)\}.$$

We can now project $v^2(+\infty, \lambda)$ and $\tilde{v}^2(+\infty, \lambda)$ onto the basis $v^1(+\infty, \lambda)$ and $\tilde{v}^1(+\infty, \lambda)$ according to (27)-(28) to obtain

$$[v^2(+\infty, \lambda), \tilde{v}^2(+\infty, \lambda)] = [\tilde{v}^1(+\infty, \lambda), v^1(+\infty, \lambda)] S, \quad (29)$$

where

$$S = \begin{pmatrix} a(\lambda) & b^*(\lambda^*) \\ b(\lambda) & -a^*(\lambda^*) \end{pmatrix}. \quad (30)$$

The matrix S is called the scattering matrix and contains the nonlinear Fourier coefficients [11], [12]. It is a function of $q(t, \cdot)$ and says how the solution to (21) is scattered from $t = -\infty$ to $t = +\infty$. More precisely, the field $v^2(-\infty, \lambda) = [1, 0]^T e^{-j\lambda t}$ is applied at $t = -\infty$, where q is absent. This field evolves forward in time according to (21), interacts with the signal (which can be viewed as an “obstacle”) at finite values of t , and subsequently propagates towards $t = +\infty$, where again q is absent. The field at $t = +\infty$ is measured and gives information about the “obstacle” as seen from a distance. Although not obvious from the development so far, we shall see in Section VII that the information measured at $t = +\infty$, captured by $a(\lambda)$ and $b(\lambda)$, is *complete*, in the sense that from this information we can retrieve $q(t, \cdot)$ entirely. In view of this interpretation, the nonlinear Fourier transform was historically referred to as the *inverse scattering transform*.

B. The Nonlinear Fourier Transform

The projection equations (27) and (28) that give $a(\lambda)$ and $b(\lambda)$ are well-defined if $\lambda \in \mathbb{R}$. From Lemma 2, Property 1), we observe that the eigenspace is symmetric in λ , i.e., if λ is an eigenvalue then so is λ^* . Thus it is sufficient to consider the upper half complex plane $\mathbb{C}^+ = \{\lambda | \Im(\lambda) > 0\}$. In this region, the boundary conditions on the basis vectors v^1 and \tilde{v}^1 at $t = \infty$ decay and blow up, respectively. As a result, (27) is consistent in \mathbb{C}^+ only if $a(\lambda) = 0$. Eigenvalues in \mathbb{C}^+ are therefore identified as the zeros of the complex function $a(\lambda)$ and they form the discrete (point) spectrum of the signal. We will see in the next section that the discrete spectrum corresponds to soliton pulses.

Lemma 4. *If $q(t) \in L_1(\mathbb{R})$, $a(\lambda)$ is an analytic function of λ on \mathbb{C}^+ .*

Proof: From Lemma 5 in Section VII, if $q(t) \in L_1(\mathbb{R})$ the scaled canonical eigenvectors $v^1(t, \lambda)e^{-j\lambda t}$ and $v^2(t, \lambda)e^{j\lambda t}$ are analytic functions of λ in \mathbb{C}^+ . Since

$$a(\lambda) = \langle v^2, v^1 \rangle_s = \langle v^2 e^{j\lambda t}, v^1 e^{-j\lambda t} \rangle_s,$$

is a combination of two analytic functions in \mathbb{C}^+ , it is analytic in the same region. (Note, however, that $b(\lambda)$, which is a combination of functions analytic in disjoint regions in \mathbb{C} , may not be analytic in either of those regions.) \square

A consequence of Lemma 4 is that the zeros of $a(\lambda)$ in \mathbb{C}^+ are isolated points [24]. It follows that the Zakharov-Shabat operator for the NLS equation has two types of spectra. The discrete (or point) spectrum, which occurs in \mathbb{C}^+ , is characterized by those $\lambda_j \in \mathbb{C}^+$ satisfying

$$a(\lambda_j) = 0, \quad j = 1, 2, \dots, N.$$

The discrete spectrum corresponds to solitons, and in this case (27) reduces to

$$v^2(t, \lambda_j) = b(\lambda_j)v^1(t, \lambda_j).$$

The continuous spectrum, which in general includes the whole real line $\Im(\lambda) = 0$, corresponds to the non-solitonic (or radiation) component of the signal. The continuous spectrum is the component of the NFT which corresponds to the ordinary Fourier transform, whereas the discrete spectrum has no analogue in linear systems theory. The reader is referred to Appendix A for a number of examples illustrating various notions of the spectrum associated with bounded linear operators; Example 9 is particularly relevant to the discussion here.

To distinguish between the discrete and continuous spectra, we find it convenient to refer to discrete spectral values of λ using the symbol λ_j (with a subscript). Continuous spectral values are denoted as λ (without a subscript). In general, $\lambda_j \in \mathbb{C}^+$, whereas $\lambda \in \mathbb{R}$.

For the purpose of developing the inverse transform, we find it sufficient to work with the ratios

$$\hat{q}(\lambda) = \frac{b(\lambda)}{a(\lambda)}, \quad \tilde{q}(\lambda_j) = \frac{b(\lambda_j)}{a'(\lambda_j)},$$

where $a'(\lambda_j)$ denotes the derivative $\left. \frac{da(\lambda)}{d\lambda} \right|_{\lambda=\lambda_j}$.

We can now formally define the nonlinear Fourier transform of a signal with respect to a given Lax operator L as follows.

Definition 4 (Nonlinear Fourier transform [11], [13]). Let $q(t)$ be a sufficiently smooth function in $L_1(\mathbb{R})$. The nonlinear Fourier transform of $q(t)$ with respect to a given Lax operator L consists of the continuous and discrete spectral functions $\hat{q}(\lambda) : \mathbb{R} \mapsto \mathbb{C}$ and $\tilde{q}(\lambda_j) : \mathbb{C}^+ \mapsto \mathbb{C}$ where

$$\hat{q}(\lambda) = \frac{b(\lambda)}{a(\lambda)}, \quad \tilde{q}(\lambda_j) = \frac{b(\lambda_j)}{a'(\lambda_j)}, \quad j = 1, 2, \dots, N,$$

in which λ_j are the zeros of $a(\lambda)$. Here, the spectral coefficients $a(\lambda)$ and $b(\lambda)$ are given by

$$a(\lambda) = \lim_{t \rightarrow \infty} v_1^2 e^{j\lambda t}, \quad (31)$$

$$b(\lambda) = \lim_{t \rightarrow \infty} v_2^2 e^{-j\lambda t}, \quad (32)$$

where $v^2(t, \lambda)$ is a solution of (21) under the boundary condition (25). \square

To obtain the continuous spectral function $\hat{q}(\lambda)$, $\lambda \in \mathbb{R}$, it is not necessary to obtain $a(\lambda)$ and $b(\lambda)$ separately. For

convenience, one can instead write an explicit differential equation

$$\frac{dy(t, \lambda)}{dt} + q(t)e^{2j\lambda t}y^2(t, \lambda) + q^*(t)e^{-2j\lambda t} = 0, \\ y(-\infty, \lambda) = 0, \quad (33)$$

and obtain $\hat{q}(\lambda) = \lim_{t \rightarrow \infty} y(t, \lambda)$.

Analogously, one can solve the second-order differential equation

$$\frac{d^2 z(t, \lambda)}{dt^2} - \left(2j\lambda + \frac{q_t}{q}\right) \frac{dz(t, \lambda)}{dt} + |q|^2 z(t, \lambda) = 0, \\ z(-\infty, \lambda) = 1, \quad \frac{dz(-\infty, \lambda)}{dt} = 0, \quad (34)$$

and obtain $a(\lambda) = \lim_{t \rightarrow \infty} z(t, \lambda)$. The discrete spectrum is obtained as the zeros of $a(\lambda)$.

Just like the ordinary Fourier transform, the nonlinear Fourier transform can be computed analytically only in a few cases. An example is given in the next subsection.

C. Example: Nonlinear Fourier Transform of a Rectangular Pulse

Consider the rectangular pulse

$$q(t) = \begin{cases} A, & t \in [t_1, t_2]; \\ 0, & \text{otherwise.} \end{cases}$$

Let $T = t_2 - t_1$ and $T' = t_2 + t_1$.

In this case $P(\lambda, q)$ is time-independent when $t \in [t_1, t_2]$, and (21) under the boundary condition (25) can be easily solved in closed form. The canonical eigenvector v^2 is given by

$$v^2(t, \lambda) = \exp[(t - t_1)P]v^2(t_1, \lambda), \quad v^2(t_1, \lambda) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-j\lambda t_1}.$$

It follows that

$$v^2(\infty, \lambda) = v^2(t_2, \lambda) = \exp(PT)v^2(t_1, \lambda),$$

where

$$\exp(PT) = \exp \left\{ \begin{bmatrix} -j\lambda & q \\ -q^* & j\lambda \end{bmatrix} T \right\} \\ = \begin{bmatrix} \cos(\Delta T) - j\frac{\lambda}{\Delta} \sin(\Delta T) & \frac{A}{\Delta} \sin(\Delta T) \\ \frac{-A^*}{\Delta} \sin(\Delta T) & \cos(\Delta T) + j\frac{\lambda}{\Delta} \sin(\Delta T) \end{bmatrix},$$

with $\Delta = \sqrt{\lambda^2 + |A|^2}$. The spectral coefficients are obtained from (31) and (32) as

$$a(\lambda) = \left(\cos(\Delta T) - j\frac{\lambda}{\Delta} \sin(\Delta T) \right) e^{j\lambda T}, \\ b(\lambda) = \frac{-A^*}{\Delta} \sin(\Delta T) e^{-j\lambda T'}.$$

The zeros of $a(\lambda)$ in \mathbb{C}^+ , which satisfy

$$j \tan(T\sqrt{|A|^2 + \lambda^2}) = \sqrt{1 + \frac{|A|^2}{\lambda^2}},$$

give rise to the discrete spectrum. The continuous spectrum is given by

$$\hat{q}(\lambda) = \frac{A^*}{j\lambda} e^{-2j\lambda t_2} \left(1 - \frac{\Delta}{j\lambda} \cot(\Delta T) \right)^{-1}.$$

Note that as $A \rightarrow 0$, $\Delta \rightarrow \lambda$, and one can see that in the limit of $AT \ll 1$ there is no discrete spectrum. Furthermore, the continuous spectrum tends to

$$\hat{q}(\lambda) = -A^* T e^{-j\lambda T'} \text{sinc}(2Tf), \quad \lambda = 2\pi f,$$

which is just the ordinary Fourier transform of the $q(t)$.

Fig. 4 shows the two spectra for $T = 1$ and various values of A . For small A , there is no discrete spectrum and the continuous spectrum is essentially just the ordinary Fourier transform of $q(t)$. As A is increased, the continuous spectrum deviates from the ordinary Fourier transform and one or more discrete mass points appear on the $j\omega$ axis.

D. Elementary Properties of the Nonlinear Fourier Transform

Let $q(t) \leftrightarrow (\hat{q}(\lambda), \tilde{q}(\lambda_k))$ be a nonlinear Fourier transform pair. The following properties are proved in Appendix B.

- 1) (The ordinary Fourier transform as limit of the nonlinear Fourier transform): If $\|q\|_{L_1} \ll 1$, there is no discrete spectrum and $\hat{q}(\lambda) \rightarrow Q(\lambda)$, where $Q(\lambda)$ is the ordinary (linear) Fourier transform of $-q^*(t)$

$$Q(\lambda) = - \int_{-\infty}^{\infty} q^*(t) e^{-2j\lambda t} dt.$$

- 2) (Weak nonlinearity): If $|a| \ll 1$, then $\widehat{a\tilde{q}}(\lambda) \approx a\hat{q}(\lambda)$ and $\widetilde{a\tilde{q}}(\lambda_k) \approx a\tilde{q}(\lambda_k)$. In general, however, $\widehat{a\tilde{q}}(\lambda) \neq a\hat{q}(\lambda)$ and $\widetilde{a\tilde{q}}(\lambda_k) \neq a\tilde{q}(\lambda_k)$.
- 3) (Constant phase change): $\widehat{e^{j\phi}q}(t)(\lambda) = e^{j\phi}\widehat{q}(t)(\lambda)$ and $\widetilde{e^{j\phi}q}(t)(\lambda_k) = e^{j\phi}\widetilde{q}(t)(\lambda_k)$.
- 4) (Time dilation): $q(\frac{t}{a}) = \widetilde{a|q(a\lambda)}$ and $\widetilde{q(\frac{t}{a})} = \widetilde{a|q(a\lambda_k)}$;
- 5) (Time shift): $q(t - t_0) \leftrightarrow e^{-2j\lambda t_0} (\hat{q}(\lambda), \tilde{q}(\lambda_k))$;
- 6) (Frequency shift): $q(t)e^{-2j\omega t} \leftrightarrow (\hat{q}(\lambda - \omega), \tilde{q}(\lambda_k - \omega))$;
- 7) (Lax convolution): If $q_2(t) = q_1(t) * (L, M; \mathcal{L})$, then $\hat{q}_2(\lambda) = H(\lambda, \mathcal{L})\hat{q}_1(\lambda)$ and $\tilde{q}_2(\lambda_k) = H(\lambda, \mathcal{L})\tilde{q}_1(\lambda_k)$. For the NLS equation, the channel filter is $H(\lambda, \mathcal{L}) = \exp(-4j\lambda^2 \mathcal{L})$.

- 8) (Parseval identity): $\int_{-\infty}^{\infty} \|q(t)\|^2 dt = \hat{E} + \tilde{E}$, where

$$\hat{E} = \frac{1}{\pi} \int_{-\infty}^{\infty} \log(1 + |\hat{q}(\lambda)|^2) d\lambda, \quad \tilde{E} = 4 \sum_{j=1}^N \Im(\lambda_j).$$

The quantities \hat{E} and \tilde{E} represent the energy contained in the continuous and discrete spectra, respectively.

In addition, we have the following properties related to the nonlinear Fourier transform

- 1) (Causality and layer-peeling property): Let $q_1(t)$ and $q_2(t)$ be two signals with non-overlapping support, e.g., signals in a pulse train. Without loss of generality, assume that $q_1(t)$ is supported on $t \leq t_0$, and that $q_2(t)$ is supported on $t > t_0$. If (a_1, b_1) and (a_2, b_2)

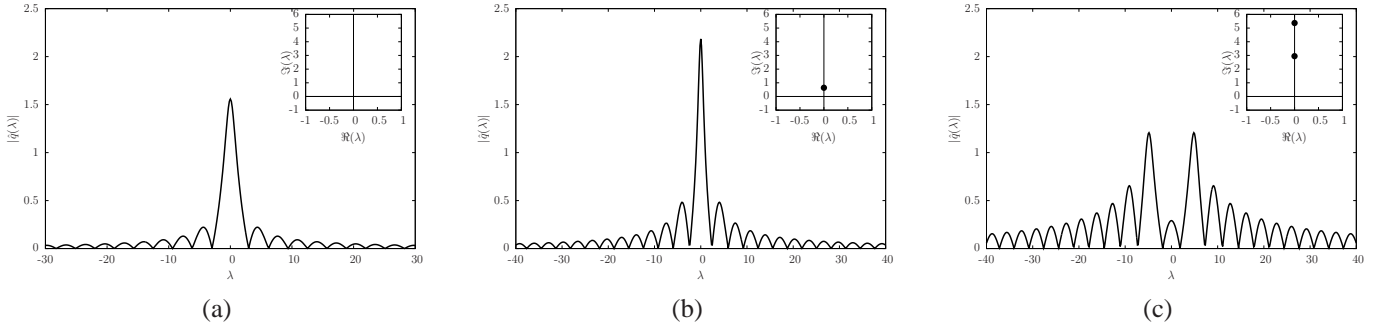


Fig. 4. Discrete and continuous spectra of the square wave signal with $T = 1$ and (a) $A = 1$ (b) $A = 2$ (c) $A = 6$.

are, respectively, the nonlinear Fourier coefficients of $q_1(t)$ and $q_2(t)$, then the nonlinear Fourier coefficients of $q_1(t) + q_2(t)$ are given by

$$\begin{aligned} (a(\lambda), b(\lambda)) &= (a_1(\lambda), b_1(\lambda)) \circ (a_2(\lambda), b_2(\lambda)) \\ &\triangleq (a_1(\lambda)a_2(\lambda) - b_1(\lambda)b_2^*(\lambda^*), a_1(\lambda)b_2(\lambda) + b_1(\lambda)a_2^*(\lambda^*)) \end{aligned}$$

That is to say, if we slice the signal in time in consecutive portions according to a mesh $-\infty < \dots < t_{-1} < t_0 < t_1 < \dots < \infty$, the nonlinear Fourier coefficients satisfy the Markov property

$$\{a_k(\lambda), b_k(\lambda); q(t), t_k < t < t_{k+1}\} \rightarrow (a_{k+1}(\lambda), b_{k+1}(\lambda)),$$

where $(a_k(\lambda), b_k(\lambda))$ are the spectral coefficients calculated from $q(t)$ in $-\infty < t < t_k$.

- 2) (Purely imaginary eigenvalues) If $\|q\|_{L_1} < \frac{\pi}{2}$, then there is no discrete spectrum. If $q(t)$ is real, nonnegative, piecewise smooth, and single-lobe (non-decreasing for $t < t_0$ and non-increasing for $t > t_0$), then there are precisely $N = \lfloor \frac{1}{2} + \frac{\|q(t)\|_{L_1}}{\pi} - \epsilon \rfloor$ eigenvalues, all purely imaginary and simple [25].

V. EVOLUTION OF THE NONLINEAR FOURIER TRANSFORM

Derivation of the evolution of the nonlinear Fourier transform of a signal propagating based on the NLS equation proceeds straightforwardly. As $q(t, z)$ propagates, the eigenvalues of L are preserved and the eigenvectors of L propagate based on (22). Assuming that $q(t, z)$ and its time-derivative vanish at $t = \pm\infty$ for all $z \leq \mathcal{L}$ during the propagation, then as $t \rightarrow \infty$ (22) is reduced to

$$v_z(t, z) \rightarrow \begin{pmatrix} 2j\lambda^2 & 0 \\ 0 & -2j\lambda^2 \end{pmatrix} v(t, z). \quad (35)$$

Thus the boundary conditions (24) and (25) are transformed to

$$\begin{aligned} v^1(t, \lambda) &\rightarrow \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{j\lambda t} e^{-2j\lambda^2 z}, \quad t \rightarrow +\infty, \\ v^2(t, \lambda) &\rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-j\lambda t} e^{2j\lambda^2 z}, \quad t \rightarrow -\infty. \end{aligned}$$

These transformed boundary conditions are not consistent with the boundary conditions (24) and (25) used to define the canonical eigenvectors, due to the additional factors $e^{\pm 2j\lambda^2 z}$. As a result, the evolution of the canonical eigenvectors from $z = 0$, according to $v_z = Mv$, does not lead to the canonical

eigenvectors at $z = \mathcal{L}$. However, by proper scaling, one can obtain the canonical eigenvectors at any z .

For instance, focusing on $v^2(t, \lambda; z)$, and changing variables to $u^2(t, \lambda; z) = v^2(t, \lambda; z)e^{-2j\lambda^2 z}$, we obtain $u_t^2 = Pu^2$ with boundary condition (25) for all z . Consequently, $u^2(t, \lambda; z)$ is a canonical eigenvector for all z . By transforming (35), the evolution equation for u is asymptotically (at $t = +\infty$) given by

$$u_z^2(\infty, \lambda; z) = \begin{pmatrix} 0 & 0 \\ 0 & -4j\lambda^2 z \end{pmatrix} u^2(\infty, \lambda; z),$$

which gives

$$\begin{aligned} u_1^2(\infty, \lambda; z) &= u_1^2(\infty, \lambda; 0) \\ u_2^2(\infty, \lambda; z) &= u_2^2(\infty, \lambda; 0)e^{-4j\lambda^2 z} \end{aligned}$$

Using expressions (31) and (32) we obtain

$$\begin{aligned} a(\lambda, z) &= \lim_{t \rightarrow \infty} u_1^2(t, \lambda; z)e^{j\lambda t} = \lim_{t \rightarrow \infty} u_1^2(t, \lambda; 0)e^{j\lambda t} = a(\lambda, 0) \\ b(\lambda, z) &= \lim_{t \rightarrow \infty} u_2^2(t, \lambda; z)e^{-j\lambda t} = \lim_{t \rightarrow \infty} u_2^2(t, \lambda; 0)e^{-4j\lambda^2 z}e^{-j\lambda t} \\ &= b(\lambda, 0)e^{-4j\lambda^2 z}. \end{aligned}$$

In turn, the nonlinear Fourier transform propagates according to

$$\begin{aligned} \widehat{q(t, z)}(\lambda) &= e^{-4j\lambda^2 z} \widehat{q(t, 0)}(\lambda), \\ \widehat{q(t, z)}(\lambda_j) &= e^{-4j\lambda_j^2 z} \widehat{q(t, 0)}(\lambda_j), \\ \lambda_j(z) &= \lambda_j(0), \quad j = 1, 2, \dots, N. \end{aligned} \quad (36)$$

Note that, since $a(\lambda, z)$ is preserved under the evolution (i.e., is independent of z), the number of the discrete eigenvalues—which are zeros of $a(\lambda)$ —are also preserved.

In summary, we see that the operation of the Lax convolution in the nonlinear Fourier domain is described by a simple multiplicative (diagonal) operator, much in the same way that the ordinary Fourier transform maps $y(t) = x(t) * h(t)$ to $Y(\omega) = X(\omega) \cdot H(\omega)$. The channel filter in (36) is $\exp(-4j\lambda^2 z)$.

VI. AN APPROACH TO COMMUNICATION OVER INTEGRABLE CHANNELS

Since the nonlinear Fourier transform of a signal is essentially preserved under Lax convolution, one can immediately conceive of a nonlinear analogue of orthogonal frequency division multiplexing (OFDM) for communication over integrable

channels. We refer to this scheme as nonlinear frequency division multiplexing (NFDM). In this scheme, the input output channel model is given by

$$\begin{aligned}\hat{Y}(\lambda) &= H(\lambda)\hat{X}(\lambda) + \hat{Z}(\lambda) \\ \tilde{Y}(\lambda_j) &= H(\lambda_j)\tilde{X}(\lambda_j) + \tilde{Z}(\lambda_j),\end{aligned}$$

where $\hat{X}(\lambda) = \hat{q}(\lambda, 0)$ and $\tilde{X}(\lambda_j) = \tilde{q}(\lambda_j, 0)$ are spectra at the input of the channel, $\hat{Y}(\lambda) = \hat{q}(\lambda, z)$ and $\tilde{Y}(\lambda_j) = \tilde{q}(\lambda_j, z)$ are spectra at the output of the channel, $\hat{Z}(\lambda)$ and $\tilde{Z}(\lambda_j)$ are effective noises in the spectral domain, and the *channel filter* is

$$H(\lambda) = e^{-4j\lambda^2 z}.$$

A bandlimited Gaussian noise in time domain generally maps to non-Gaussian noise processes in the spectral domain.

The proposed scheme consists of two steps.

- *The inverse nonlinear Fourier transform at the transmitter (INFT).* At the transmitter, information is encoded in the nonlinear spectra of the signal according to a suitable constellation on $(\hat{X}(\lambda), \tilde{X}(\lambda_j))$. The time domain signal is generated by taking the inverse nonlinear Fourier transform,

$$q(t) = \text{INFT}(\hat{X}(\lambda), \tilde{X}(\lambda_j)),$$

and is sent over the channel. (The INFT is described formally in the next section.)

- *The forward nonlinear Fourier transform at the receiver (NFT).* At the receiver, the (forward) nonlinear Fourier transform of the signal,

$$(\hat{Y}(\lambda), \tilde{Y}(\lambda_j)) = \text{NFT}(q(t, z))$$

is taken and the resulting spectra are compared against the transmitted spectra according to some metric $d(\hat{X}(\lambda), \tilde{X}(\lambda_j); \hat{Y}(\lambda), \tilde{Y}(\lambda_j))$.

As $q(t, 0)$ propagates in the time domain based on the complicated nonlinear equation, it is significantly distorted and, among other distortions, undergoes intersymbol interference (ISI) in a very complicated manner. Despite this, the nonlinear spectra of the signal, $(\hat{X}(\lambda), \tilde{X}(\lambda_j))$, are essentially held invariant during propagation, up to a complex-valued scale factor (see (36)). All the nonlinear spectral components propagate independently of each other, and the channel is decomposed into a number of linear parallel independent channels of the form $y_k = c_k x_k + n_k$, where n_k denotes the effective noise in the k th channel. By diagonalizing the channel in this way, the ISI is removed in the spectral domain.

In this scheme, as in linear OFDM, communication objectives, such as constellation design, coding and modulation are entirely formulated in the spectral domain. All available degrees of freedom, i.e., $\{\lambda_j, \tilde{X}(\lambda_j), \hat{X}(\lambda)\}$ can be generally modulated. Time domain constraints can be translated to constraints in the spectral domain. A power constraint, for instance, can be exactly transformed to a power constraint in the spectral domain with the help of the Parseval Identity. A bandwidth limitation is not directly and simply expressed in the nonlinear spectral domain; however, it appears that the

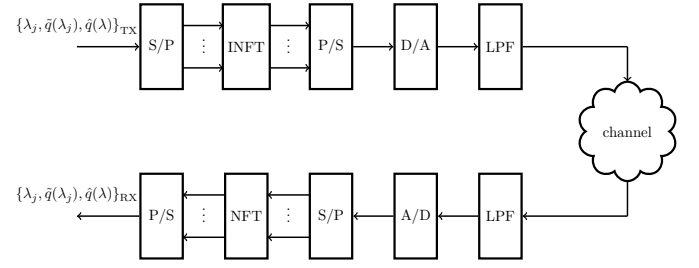


Fig. 5. Nonlinear frequency division channel model

nonlinear spectrum of a signal bandlimited to W is indeed confined, approximately, to a vertical strip in the real line and the upper half complex plane, with a width proportional to the W .

Fig. 5 illustrates the NFDM channel model that we use in this paper (and its subsequent parts) for data transmission over integrable channels such as optical fibers.

VII. INVERSE NONLINEAR FOURIER TRANSFORM

In this section, we describe the inverse nonlinear Fourier transform, which is a method for recovering the signal $q(t, \cdot)$ from its nonlinear Fourier transform $(\hat{q}(\lambda), \tilde{q}(\lambda_j))$. This is the opposite of what was done in the Section IV, and it gives further important insight into the mechanism of the NFT. As in Section IV, the value of z is irrelevant here and the index z is thus suppressed.

A. Riemann-Hilbert Factorization

In the 1950s, before the publication of Gardner [16], a method for retrieving the potential $q(t)$ in the linear Schrödinger equation from the knowledge of the scattering matrix S was already known in quantum mechanics. It was later realized that this method is an instance of a *Riemann-Hilbert problem* in complex analysis [12], [13]. Alternatively, the same task can be accomplished using the *Gelfand-Levitan method* [12], a different approach developed earlier in the context of inverse problems for Sturm-Liouville differential equations. Although either of these methods can be used for solving the inverse problem, in this paper we will use the Riemann-Hilbert method.

To begin, canonical eigenvectors can be found in two ways. On the one hand, they are related to $q(t)$ through the Zakharov-Shabat system (21). One can solve (21) to explicitly express the canonical eigenvectors as a series involving $q(t)$. On the other hand, canonical eigenvectors are related to the nonlinear Fourier transform via the projection equations (27)-(28). The latter are two equations for four unknowns and in general cannot be solved. However, the unique properties of the canonical eigenvectors, which are analytic in disjoint regions of the complex plane, will allow us to find them in terms of the nonlinear Fourier transform. Note that this second derivation does not explicitly depend on $q(t)$; however, by equating the canonical eigenvectors obtained from these two derivations, we can relate the signal $q(t)$ to its nonlinear Fourier transform $(\hat{q}(\lambda), \tilde{q}(\lambda_j))$.

We find it helpful to briefly introduce the tool that we employ in this section, namely the Riemann-Hilbert factorization problem in complex analysis.

Definition 5 (Riemann-Hilbert factorization). The Riemann-Hilbert factorization problem consists of finding two complex functions $f^-(z)$ and $f^+(z)$, respectively, analytic inside and outside of a closed contour C in the complex plane, such that on C they satisfy the boundary condition $f^+(z) = g(z)f^-(z)$ for some given $g(z)$.

It $g(z)$ satisfies a Hölder condition, it is possible to find both $f^+(z)$ and $f^-(z)$ everywhere in the complex plane from the given boundary condition. In Appendix C, we provide a brief overview of this problem and its solution. See [26] for more discussion.

B. The Inverse Transform

The inverse transform maps the two spectral functions $(\hat{q}(\lambda), \tilde{q}(\lambda_j))$ to the signal $q(t)$.

As noted, we need to express the canonical eigenvectors in terms of the nonlinear Fourier transform. Below we will find it more convenient to work with canonical eigenvectors subject to fixed boundary conditions. Scaling the canonical eigenvectors as $V^2 = v^2 e^{j\lambda t}$, $\tilde{V}^2 = \tilde{v}^2 e^{-j\lambda t}$, $V^1 = v^1 e^{-j\lambda t}$ and $\tilde{V}^1 = \tilde{v}^1 e^{j\lambda t}$, the projection equations (27) and (28) are transformed to

$$V^2(t, \lambda) = a(\lambda)\tilde{V}^1(t, \lambda) + b(\lambda)e^{2j\lambda t}V^1, \quad (37)$$

$$\tilde{V}^2(t, \lambda^*) = -a^*(\lambda^*)V^1(t, \lambda) + b^*(\lambda^*)e^{-2j\lambda t}\tilde{V}^1. \quad (38)$$

Lemma 5. If $q \in L^1(\mathbb{R})$, then $V^1(t, \lambda)$ and $V^2(t, \lambda)$ are analytic functions of λ in the upper half complex plane $\mathbb{C}^+ = \{z \in \mathbb{C} : \Im(z) > 0\}$ while $\tilde{V}^1(t, \lambda)$ and $\tilde{V}^2(t, \lambda)$ are analytic functions of λ in the lower half complex plane $\mathbb{C}^- = \{z \in \mathbb{C} : \Im(z) < 0\}$.

Proof: See Appendix D. \square

Rearranging (37)-(38) gives

$$\begin{pmatrix} V^1 & V^2 \end{pmatrix} = \begin{pmatrix} \tilde{V}^1 & \tilde{V}^2 \end{pmatrix} \begin{pmatrix} \frac{b^*(\lambda^*)}{a^*(\lambda^*)}e^{-2j\lambda t} & \frac{1}{a^*(\lambda^*)} \\ \frac{-1}{a^*(\lambda^*)} & -\frac{b(\lambda)}{a^*(\lambda^*)}e^{-2j\lambda t} \end{pmatrix}. \quad (39)$$

Since $(V^2 \ V^1)$ and $(\tilde{V}^2 \ \tilde{V}^1)$ are analytic, respectively, in the upper and lower half-planes, (39) defines a matrix Riemann-Hilbert problem.

Solution of the Riemann-Hilbert factorization problem (39) is given in Appendix F. Omitting the details, the following linear system of equations, known as the Riemann-Hilbert

system, is obtained:

$$\begin{aligned} \tilde{V}^1(t, \lambda) &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \sum_{i=1}^N \frac{\tilde{q}(\lambda_i)e^{2j\lambda_i t}V^1(t, \lambda_i)}{\lambda - \lambda_i} \\ &\quad + \frac{1}{2\pi j} \int_{-\infty}^{\infty} \frac{\hat{q}(\zeta)e^{2j\zeta t}V^1(t, \zeta)}{\zeta - (\lambda - j\epsilon)} d\zeta, \\ V^1(t, \lambda) &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \sum_{i=1}^N \frac{\tilde{q}^*(\lambda_i^*)e^{-2j\lambda_i^* t}\tilde{V}^1(t, \lambda_i)}{\lambda - \lambda_i^*} \\ &\quad + \frac{1}{2\pi j} \int_{-\infty}^{\infty} \frac{\hat{q}^*(\zeta^*)e^{-2j\zeta^* t}\tilde{V}^1(t, \zeta)}{\zeta - (\lambda + j\epsilon)} d\zeta, \end{aligned}$$

$$\begin{aligned} \tilde{V}^1(t, \lambda_m) &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \sum_{i=1}^N \frac{\tilde{q}(\lambda_i)e^{2j\lambda_i t}V^1(t, \lambda_i)}{\lambda_m^* - \lambda_i} \\ &\quad + \frac{1}{2\pi j} \int_{-\infty}^{\infty} \frac{\hat{q}(\zeta)e^{2j\zeta t}V^1(t, \zeta)}{\zeta - \lambda_m^*} d\zeta, \quad m = 1, \dots, N, \\ V^1(t, \lambda_m) &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \sum_{i=1}^N \frac{\tilde{q}^*(\lambda_i^*)e^{-2j\lambda_i^* t}\tilde{V}^1(t, \lambda_i)}{\lambda_m - \lambda_i^*} \\ &\quad + \frac{1}{2\pi j} \int_{-\infty}^{\infty} \frac{\hat{q}^*(\zeta^*)e^{-2j\zeta^* t}\tilde{V}^1(t, \zeta)}{\zeta - \lambda_m} d\zeta, \\ m &= 1, \dots, N. \end{aligned} \quad (40)$$

This is a system of $2N + 2$ linear equations for $2N + 2$ discrete and continuous canonical eigenvectors $\{V^1(t, \lambda_m)\}_{m=1}^N$, $\{\tilde{V}^1(t, \lambda_m)\}_{m=1}^N$, and $V^1(t, \lambda)$ and $\tilde{V}^1(t, \lambda)$ as a function of the Fourier transform $(\hat{q}(\lambda), \tilde{q}(\lambda_j))$.

Next, the canonical eigenvectors are related to the signal $q(t)$. By inspecting the Zakharov-Shabat system (21), it is shown in Appendix E that for $|\lambda| \gg 1$

$$V^1(t, \lambda) = \begin{pmatrix} \frac{1}{2j\lambda}q(t) \\ 1 + \frac{1}{2j\lambda} \int_{-\infty}^{\infty} |q(t)|^2 dt \end{pmatrix} + O(\lambda^{-2}). \quad (41)$$

It now becomes easy to recover the signal $q(t)$ from the nonlinear Fourier transform $(\hat{q}(\lambda), \tilde{q}(\lambda_j))$. Eigenvector V^1 is related to q via (41) for $|\lambda| \gg 1$, and is related to the nonlinear Fourier transform via the second equation in (40). Approximating $1/(\lambda - \zeta)$ and $1/(\lambda - \lambda_i)$ by $-1/\lambda + O(\lambda^{-2})$ in the second equation of (40) for $|\lambda| \gg 1$ and comparing V_1^1 with V_1^1 in (41), we obtain

$$\begin{aligned} q^*(t) &= 2j \sum_{i=1}^N \tilde{q}(\lambda_i)e^{2j\lambda_i t}V_2^1(t, \lambda_i) \\ &\quad - \frac{1}{\pi} \int_{-\infty}^{\infty} \hat{q}(\lambda)e^{2j\lambda t}V_2^1(t, \lambda)d\lambda. \end{aligned} \quad (42)$$

This represents $q(t)$ as a function of the nonlinear Fourier transform.

In summary, given $(\hat{q}(\lambda), \tilde{q}(\lambda_j))$, we first solve the Riemann-Hilbert system to find the eigenvector V^1 . For this

purpose, one can discretize the system (40) and solve a linear system of equations of the form

$$Ax = b,$$

for appropriate A and b . This is done for each fixed t_i to determine the samples of $V^l(t_i, \lambda)$ at that time. Then, V^1 , \hat{q} and \tilde{q} are substituted in (42) to obtain the signal $q(t)$.

Note that this inverse transform is taken only once at the transmitter to synthesize the desired pulse shapes. It is only the forward transform which needs to be taken in real time at the receiver.

The mathematical tools developed in this paper are used in Parts II, III and IV of this paper to present further details pertaining the application of the suggested scheme in data transmission over integrable channels.

In Part II, numerical methods are suggested to compute the nonlinear Fourier transform of a signal with respect to the AKNS Lax operator. We use the results of Part I and II in Part III and IV to study the NFT in the presence of the noise and illustrate the performance of the method in actual fiber-optic systems.

VIII. CONCLUSIONS

The nonlinear Fourier transform of a signal with respect to an operator L in a Lax pair consists of continuous and discrete spectral functions $\hat{x}(\lambda)$ and $\hat{x}(\lambda_j)$, obtainable by solving the eigenproblem for the L operator. The NFT maps a Lax convolution to a multiplication operator in the spectral domain. Using the nonlinear Fourier transform, we propose a transmission scheme for integrable channels, termed nonlinear frequency-division multiplexing, in which the information is encoded in the nonlinear spectrum of the signal. The scheme is an extension of traditional OFDM to any channel generated by a Lax pair. An example is the optical fiber channel, in which signals propagate according to the nonlinear Schrödinger equation. The class of integrable channels, though nonlinear and often complicated, are somehow “linear in disguise,” and thus admit the proposed nonlinear frequency-division multiplexing transmission scheme.

APPENDIX A SPECTRUM OF BOUNDED LINEAR OPERATORS

When moving from finite-dimensional spaces (of e.g., matrices) to infinite-dimensional spaces (of e.g., functions and operators), some results do not carry over necessarily. Here we recall a few useful results in functional analysis [20].

Let \mathcal{H} be a Hilbert space, let \mathcal{D} be a dense subset of \mathcal{H} , and let $L : \mathcal{D} \rightarrow \mathcal{H}$ be an operator.

Definition 6. The *adjoint* of L is the operator L^* whose domain \mathcal{D}^* consists of all $\psi \in \mathcal{H}$ for which there exists an element $L^*\psi \in \mathcal{H}$ such that

$$\langle L^*\psi, \phi \rangle = \langle \psi, L\phi \rangle, \quad \forall \phi \in \mathcal{D}.$$

The operator L is said to be *self-adjoint* if $\mathcal{D} = \mathcal{D}^*$ and $L^* = L$. Ignoring domain subtleties (as is the case for bounded operators), self-adjoint operators are the analogue of Hermitian matrices.

Definition 7. Given an operator L on a Hilbert space, an operator M is said to be the inverse of L if the domain of M is the range of L , the range of M is the domain of L , and $ML = I$ and $LM = I$.

We will restrict ourselves now to operators whose \mathcal{D} is the entire space \mathcal{H} .

An operator $L : \mathcal{H} \rightarrow \mathcal{H}$ is invertible if it 1) is one-to-one 2) is onto 3) has bounded inverse. In finite-dimensional spaces, only the first condition is required.

An operator is bounded if it maps bounded inputs to bounded outputs. A bounded operator is invertible if it is one-to-one and onto.

Definition 8. The *spectrum* of an operator L on \mathcal{H} is defined as

$$\sigma(L) = \{\lambda \in \mathbb{C} | L - \lambda I \text{ is not invertible}\}.$$

The spectrum of a bounded operator can be partitioned into two classes, depending on the reason that $L - \lambda I$ fails to be invertible.

A complex number λ is considered part of the *discrete spectrum* if $L - \lambda I$ is not one-to-one, i.e., $L\psi = \lambda\psi$ has a non-zero solution $\psi \in \mathcal{H}$. In this case, λ is called an *eigenvalue*, and each ψ satisfying this equation for the given λ is an associated *eigenvector*. The set of all eigenvalues is called the *point spectrum* or *discrete spectrum* of L , $\sigma_{\text{disc}}(L)$.

It can also happen that $L - \lambda I$ fails to be surjective, i.e., the range of $L - \lambda I$ is a proper subset of \mathcal{H} . We call the set of λ for which this happens the *continuous spectrum* $\sigma_{\text{cont}}(L)$ [20].

In finite-dimensional Hilbert spaces the spectrum is entirely discrete. This may no longer be true in infinite-dimensional spaces, where the eigenvalues (if they exist) may only be one part of the spectrum. The spectrum of a self-adjoint operator is real. In general, $\sigma(L) = \sigma(L^*)$.

The following examples illustrate some these possibilities.

Example 5. The operator $L(x(t)) = tx(t)$, $x(t) \in L^2[0, 1]$, has no eigenvalues and its spectrum is purely continuous $\sigma(L) = [0, 1]$. \square

Example 6. The Fourier transform operator $\mathcal{F}(q)(\omega) = \int_{-\infty}^{\infty} q(t)e^{j\omega t} dt$, regarded as an operator on $L^2(\mathbb{R})$ has the property $\mathcal{F}^4 = I$. The eigenvalues are therefore the discrete values $\{\pm 1, \pm j\}$. If $p(t)$ is an arbitrary polynomial, then $p(t)\exp(-t^2/2)$ is an eigenfunction. \square

Example 7. Let ∇^2 be the Laplace operator and let r denote the radial distance in the three-dimensional space. The operator $L = -\nabla^2 - \frac{1}{r}$ is self-adjoint, and therefore it has a real spectrum. The continuous spectrum is $[0, \infty)$ and the discrete spectrum is given by $\lambda_n = -\frac{1}{4n^2}$, $n = 1, 2, \dots$. \square

Example 8. It is possible that the discrete spectrum of an operator is uncountable. For example, for a sequence $(x_0, x_1, \dots) \in \ell^2 = \{x : \sum_i |x_i|^2 < \infty\}$, let the left-shift operator L be defined as $L(x_0, x_1, x_2, \dots) = (x_1, x_2, \dots)$. The spectrum consists of the unit disk $|\lambda| \leq 1$. The portion $|\lambda| < 1$ is the discrete spectrum while $|\lambda| = 1$ is the continuous spectrum. The adjoint of L is the right-shift operator

$R(x_0, x_1, x_2, \dots) = (0, x_0, x_1, \dots)$. This operator has the same total spectrum, but it is entirely continuous. \square

Example 9. Consider the Zakharov-Shabat system (21), in the form $(DI - P)v = 0$, $v \in L^2(\mathbb{R})$. One can see that for $\lambda \in \mathbb{R}$, the eigenvectors $v(t, \lambda)$ contain oscillatory terms $\exp(\pm j\lambda t)$, and thus are not in $L^2(\mathbb{R})$ (unbounded in norm). As a result, the operator $(DI - P)^{-1}$ is unbounded and the inverse $(DI - P)^{-1}$ does not exist on the Hilbert space $L^2(\mathbb{R})$. Thus $\lambda \in \mathbb{R}$ forms the continuous spectrum. This is generally the case where for an operator L on \mathcal{H} for which the eigenproblem $Lv = \lambda v$ has a solution with infinite norm. \square

Note that some authors subdivide what we refer to as the continuous spectrum into further classes (e.g., residual spectrum, essential spectrum, etc.) [20]; however, for the purposes of this paper, classification into discrete and continuous spectra will suffice.

Hermitian matrices are always diagonalizable, have real eigenvalues, and possess a complete set of orthonormal eigenvectors, which provide a basis for the input space. There is a perfect generalization of this result to *compact* self-adjoint operators [20].

Theorem 6 (Hilbert-Schmidt spectral theorem). *Let L be a compact self-adjoint operator in \mathcal{H} . Then it is always possible to find eigenfunctions $\{\psi_n\}$ of L forming an orthonormal basis for \mathcal{H} .*

The spectral theory of operators that are not compact and self-adjoint is more involved; see [20].

An important class of operators are the multiplication operators, which are analogous to diagonal matrices.

Definition 9 (Multiplication operator). Let $f(t)$ be an arbitrary function. The operator L defined by $(L\psi)(t) = f(t)\psi(t)$, which performs sample-wise multiplication, is called the *multiplication operator* or *diagonal operator* induced by $f(t)$.

Theorem 7. *Every bounded self-adjoint operator in a separable Hilbert space \mathcal{H} is unitarily equivalent to a multiplication operator Γ , i.e., $L = U\Gamma U^{-1}$ where U is unitary.*

It follows that the essence of a bounded self-adjoint operator is just a multiplication operator.

APPENDIX B

PROOF OF ELEMENTARY PROPERTIES OF THE NFT

In this section, we sketch the proofs of the properties of the NFT stated in Section IV-D.

- 1) If $\|q(t)\|_{L_1} \ll 1$, then $y^2(t, \lambda)$ and $q^2(t)$ terms can be ignored in (33) and (34). From the resulting equations, it follows that there is no discrete spectrum and $\hat{q}(\lambda) \rightarrow Q(\lambda)$. The quadratic terms are introduced by the NFT to account for the nonlinearity.
- 2) This follows from 1) in above, and that when $|a| \ll 1$, the squared terms representing the nonlinearity can be ignored.
- 3) This follows from replacing $q(t)$ with $e^{j\phi}q(t)$ in (33) and (34).

- 4) Scaling the Zakharov-Shabat system (21) as $t' = t/a$, we obtain the desired result. Note that if $\text{sgn}(a) < 0$, boundary conditions at $\pm\infty$ are interchanged.
- 5) Property 5) and 6) follow by replacing $q(t)$ with $q(t-t_0)$ and $e^{2j\omega t}q(t)$ in (33) and (34), and accordingly changing variables.
- 6) This is the statement of (36).
- 7) The following identity, known as the trace formula, can be easily proved for the nonlinear Fourier transform [12]

$$c_n = \frac{4}{n} \sum_{i=1}^N \Im(\lambda_i) + \frac{1}{\pi} \int_{-\infty}^{\infty} \zeta^{n-1} \log(1 + |\rho(\zeta)|^2) d\zeta.$$

Here c_n are the secondary constants of motion, i.e., quantities, directly in terms of the time domain data, which are preserved during the flow of the NLS equation. The first few ones are the energy

$$c_1 = \int_{-\infty}^{\infty} |q(t)|^2 dt,$$

momentum

$$c_2 = \frac{1}{2j} \int_{-\infty}^{\infty} q(t)q_t^*(t) dt,$$

and the Hamiltonian

$$c_3 = - \int_{-\infty}^{\infty} (|q(t)|^4 - |q_t(t)|^2) dt.$$

Parseval's Identity is the trace formula evaluated at $n = 1$.

- 8) Let $q(t)$ be supported in the interval $[t_1, t_2]$ and let $K(q(t), t_1, t_2)$ denote a propagator (linear transformation) which maps $v(t, \lambda)$ in (21) from $v(t_1, \lambda)$ to $v(t_2, \lambda)$. The propagator is structured as

$$K = \begin{pmatrix} a(\lambda)e^{j\lambda(t_2-t_1)} & -b^*(\lambda^*)e^{j\lambda(t_2+t_1)} \\ b(\lambda)e^{-j\lambda(t_2+t_1)} & a^*(\lambda^*)e^{-j\lambda(t_2-t_1)} \end{pmatrix}.$$

Let $q_1(t)$ and $q_2(t)$, supported, respectively, in the intervals $[t_1, t_2]$ and $[t_2, t_3]$, $t_1 < t_2 < t_3$, correspond to the propagators $K_1(q_1(t), t_1, t_2)$ and $K_2(q_2(t), t_2, t_3)$. Then $q(t) = q_1(t) + q_2(t)$ is supported in $[t_1, t_3]$, and from linearity corresponds to the propagator $K = K_2K_1$

$$K = \begin{pmatrix} a_2(\lambda)e^{j\lambda(t_3-t_2)} & -b_2^*(\lambda^*)e^{j\lambda(t_3+t_2)} \\ b_2(\lambda)e^{-j\lambda(t_3+t_2)} & a_2^*(\lambda^*)e^{-j\lambda(t_3-t_2)} \end{pmatrix} \times \begin{pmatrix} a_1(\lambda)e^{j\lambda(t_2-t_1)} & -b_1^*(\lambda^*)e^{j\lambda(t_2+t_1)} \\ b_1(\lambda)e^{-j\lambda(t_2+t_1)} & a_1^*(\lambda^*)e^{-j\lambda(t_2-t_1)} \end{pmatrix}.$$

The 1×1 and 2×1 elements are

$$K_{11} = e^{j\lambda(t_3-t_1)} (a_1(\lambda)a_2(\lambda) - b_1(\lambda)b_2^*(\lambda^*)), \\ K_{21} = e^{-j\lambda(t_3+t_1)} (a_1(\lambda)b_2(\lambda) + b_1(\lambda)a_2^*(\lambda^*)).$$

Comparing with $K_{11} = a(\lambda)e^{j\lambda(t_3-t_1)}$ and $K_{21} = b(\lambda)e^{-j\lambda(t_3+t_1)}$, we get the desired result.

APPENDIX C

RIEMANN-HILBERT FACTORIZATION PROBLEM

A. Preliminary

Recall that a complex function $f(z) = u(x, y) + jv(x, y)$ in the complex plane $z = (x, y)$ is differentiable at a point (x_0, y_0) if and only if the partial derivatives u_x, u_y, v_x and v_y are continuous and

$$u_x = v_y, \quad u_y = -v_x. \quad (43)$$

The compatibility conditions (43) are called the Cauchy-Riemann conditions and are obtained by equating the limit $\delta z \rightarrow 0$ along the real and imaginary axes. This means that, unlike real functions, differentiability of a complex function imposes a constraint between the real and imaginary parts of the function.

A function $f(z)$ is said to be analytic at (x_0, y_0) if it is differentiable in a neighborhood of that point. If $f(z)$ is analytic in an open region Ω of the complex plane, then a power series representation of $f(z)$ is convergent in Ω . Existence of a power series representation has a number of interesting consequences. For instance, the zeros of a non-zero function analytic in an open region Ω of \mathbb{C} are isolated points in Ω [24].

Lemma 8 (Plemelj formulae). *Let C be any smooth, closed, counter-clockwise, contour in the complex plane and let $f(z)$ be any function satisfying a Hölder condition on C defined by*

$$|f(t) - f(\tau)| \leq k|t - \tau|^\alpha, \quad k > 0, \quad \forall t, \tau \in C,$$

for some $0 < \alpha \leq 1$. Then the projection integral

$$F(\zeta) = \frac{1}{2\pi j} \oint_C \frac{f(z)}{z - \zeta} dz, \quad (44)$$

is analytic everywhere in \mathbb{C} except possibly at points ζ on the contour C (where $F(\zeta)$ is not defined). If ζ approaches C along a path entirely inside the contour C , then

$$F^+(\zeta) \triangleq \lim_{z \rightarrow \zeta} F(\zeta) = \frac{f(\zeta)}{2} + \frac{1}{2\pi j} \text{p.v.} \int_C \frac{f(z)}{z - \zeta} dz. \quad (45)$$

If ζ approaches C along a path entirely outside the contour C , then

$$F^-(\zeta) \triangleq \lim_{z \rightarrow \zeta} F(\zeta) = -\frac{f(\zeta)}{2} + \frac{1}{2\pi j} \text{p.v.} \int_C \frac{f(z)}{z - \zeta} dz. \quad (46)$$

Here p.v. denotes the principal value integral defined by

$$\text{p.v.} \int_C \frac{f(z)}{z - \zeta} dz = \lim_{\epsilon \rightarrow 0} \oint_{C - C_\epsilon} \frac{f(z)}{z - \zeta} dz,$$

in which C_ϵ is an infinitesimal part of C centered at $z = \zeta$ and with length 2ϵ .

Proof: See [26]. ■

The projected function $F(\zeta)$ is a *sectionally analytic* function of ζ with respect to C , i.e., it is analytic in sections C^+ (the interior of C) and C^- (the exterior of C), and the limits $F^\pm(\zeta)$ exist (as given by (45) and (46)).

A consequence of Lemma 8 is that $F^\pm(\zeta)$ satisfies the following *jump condition* on the boundary C

$$F^+(t) - F^-(t) = f(t).$$

The projection operator therefore produces functions which are analytic almost everywhere, except on a contour where it experiences a jump in its limits.

1) *The scalar Riemann-Hilbert problem:* In the scalar Riemann-Hilbert problem, the task is to find functions $f^+(z)$ and $f^-(z)$, analytic, respectively, inside and outside a given smooth closed contour C , such that on C

$$f^+(t) = g(t)f^-(t) + h(t) \quad t \in C, \quad (47)$$

where $h(t)$ and $g(t)$ (with $g(t) \neq 0$ for all $t \in C$), are given functions satisfying a Hölder condition on C .

Both unknowns $f^+(z)$ and $f^-(z)$ can be determined from the single boundary equation (47), using the projection operator (44). To see this, first consider the homogeneous case where $h(t) = 0$. One can rewrite (47) as a jump condition

$$\log f^+(z) - \log f^-(z) = \log g(z) \quad z \in C.$$

Functions $\log f^+(z)$ and $\log f^-(z)$ can be viewed as portions of a single sectionally analytic function $\log f(z)$ which is analytic in C^+ and C^- and on boundary C its limits jump as $\log g(t)$. In view of the projection operator P (44), consider

$$\log f(z) = \frac{1}{2\pi j} \oint_C \frac{\log g(\lambda)}{\lambda - z} d\lambda.$$

If $\log g(t)$ satisfies a Hölder condition on C , then $\log f(z)$ is analytic strictly inside and outside C . On C , we can define $\log f^+(z)$ and $\log f^-(z)$, respectively, as equal to the limits (45) and (46). The function obtained in this way satisfies (47) and has the desired analyticity properties.

Note however that, unlike $g(t)$, $\log g(t)$ in the integrand may not satisfy a Hölder condition. To resolve this issue, we can multiply $g(t)$ by a decaying factor t^{-k} , for a suitable k , to make $t^{-k}g(t)$ Hölder, and obtain $f^+(t) = (t^{-k}g(t))(t^k f^-(z))$. Therefore, defining

$$F(z) = \exp \left(\frac{1}{2\pi j} \oint_C \frac{\log \lambda^{-k} g(\lambda)}{\lambda - z} d\lambda \right),$$

we have the following solution for the homogeneous RH problem:

$$f^+(z) = \begin{cases} F(z) & z \in C^+, \\ F^+(z) & z \in C, \end{cases}$$

and

$$f^-(z) = \begin{cases} z^{-k} F(z) & z \in C^-, \\ z^{-k} F^-(z) & z \in C. \end{cases}$$

Here k can be chosen so that $t^{-k}g(t)$ is continuous and the total phase change of $\log t^{-k}g(t)$ is zero along the closed path C .

The solution $f^\pm(z)$ is called the fundamental solution to the scalar RH problem. From the homogeneity of (47), one

can obtain other solutions by multiplying $f^\pm(z)$ by any entire function in \mathbb{C} .

We can generalize the above procedure to solve the non-homogeneous Riemann-Hilbert problem (47). In this case, we can find a factorization $g(t) = \frac{g^+(t)}{g^-(t)}$ by solving a homogeneous Riemann-Hilbert problem with boundary conditions $g^+(t) = g(t)g^-(t)$. Then (47) is reduced to

$$\frac{f^+(t)}{g^+(t)} - \frac{f^-(t)}{g^-(t)} = \frac{h(t)}{g^+(t)},$$

which, as before, can be solved in closed form using the Plemelj formulae.

2) *The matrix Riemann-Hilbert problem:* When formulating the inverse nonlinear Fourier transform, we face a matrix Riemann-Hilbert problem (39). Matrix RH problems are generally more involved and may not allow closed-form solutions [26]. As we will see in the Appendix F, for the particular matrix RH problem (39), the projection operator (44) is sufficient to solve the problem.

APPENDIX D PROOF OF LEMMA 5

Analyticity of the canonical eigenvectors is directly a property of the Zakharov-Shabat system (21). Let us scale canonical eigenvectors as $V^2 = v^2 e^{j\lambda t}$, $\tilde{V}^2 = \tilde{v}^2 e^{-j\lambda t}$, $V^1 = v^1 e^{-j\lambda t}$ and $\tilde{V}^1 = \tilde{v}^1 e^{j\lambda t}$. Transforming the Zakharov-Shabat system (21), the scaled eigenvectors satisfy

$$\begin{aligned} V_t^2 &= \begin{pmatrix} 0 & q \\ -q^* & 2j\lambda \end{pmatrix} V^2, & V^2(-\infty) &= \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\ \tilde{V}_t^2 &= \begin{pmatrix} -2j\lambda & q \\ -q^* & 0 \end{pmatrix} \tilde{V}^2, & \tilde{V}^2(-\infty) &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\ \tilde{V}_t^1 &= \begin{pmatrix} 0 & q \\ -q^* & 2j\lambda \end{pmatrix} \tilde{V}^1, & \tilde{V}^1(\infty) &= \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\ V_t^1 &= \begin{pmatrix} -2j\lambda & q \\ -q^* & 0 \end{pmatrix} V^1, & V^1(\infty) &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{aligned} \quad (48)$$

Let us, for instance, solve for the canonical eigenvector V^2 in (48). Considering the q terms as an external force and using the Duhamel formula [27], (48) can be transformed into its integral representation

$$\begin{aligned} V^2(t, \lambda) &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &+ \int_{-\infty}^{\infty} h(t-t', \lambda) \begin{pmatrix} 0 & q(t') \\ -q^*(t') & 0 \end{pmatrix} V^2(t', \lambda) dt', \end{aligned} \quad (49)$$

where the system impulse response $h(t, \lambda)$ is

$$h(t, \lambda) = \begin{pmatrix} u(t) & 0 \\ 0 & e^{2j\lambda t} u(t) \end{pmatrix}, \quad (50)$$

and we have ignored transient terms since the boundary condition starts at $t = -\infty$.

The analyticity of eigenvectors can be seen intuitively at this stage. The impulse response (50) involves the term $e^{2j\lambda t} u(t)$ and hence (48) is well defined in \mathbb{C}^+ if $q \in L^1(\mathbb{R})$. The impulse response for the V^1 equation involves $-e^{-2j\lambda t} u(-t)$ and hence it is bounded in the same region.

The impulse response for \tilde{V}^2 and \tilde{V}^1 have terms proportional to $e^{-2j\lambda t} u(t)$ and $-e^{2j\lambda t} u(-t)$, respectively, and therefore these eigenvectors are analytic in \mathbb{C}^- for $q \in L^1(\mathbb{R})$.

A more precise argument proceeds by solving (49) explicitly. Duhamel's integral (49) is of the form of a fixed-point map

$$V^2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + T(V^2),$$

where T is the linear operator underlying (49). A candidate for the solution is the sum

$$V^2 = \sum_{k=0}^{\infty} U^k \quad (51)$$

where U^k satisfy the iteration

$$U^{k+1} = T(U^k) + \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad U_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The first few terms are

$$U^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad U^2 = \begin{pmatrix} 0 \\ -\int_{t_1=-\infty}^t q^*(t_1) e^{2j\lambda(t-t_1)} dt_1 \end{pmatrix}, \quad (52)$$

$$U^3 = \begin{pmatrix} -\int_{t_2=-\infty}^t \int_{t_1=-\infty}^{t_2} q(t_2) q^*(t_1) e^{2j\lambda(t_2-t_1)} dt_1 dt_2 \\ 0 \end{pmatrix}, \quad (53)$$

and the k^{th} term is recursively defined by

$$U^{k+1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \int_{-\infty}^t \begin{pmatrix} q(t') U_2^k(t', \lambda) \\ -q^*(t') U_1^k(t', \lambda) e^{2j\lambda(t-t')} \end{pmatrix} dt'. \quad (54)$$

By induction on k , as (54) suggests, if U^k is analytic and $q(t) \in L^1(\mathbb{R})$, then U^{k+1} is analytic. In addition, the series (51) is uniformly convergent on t and thus V^2 is analytic in \mathbb{C}^+ .

Similarly one proves the analyticity of the other canonical eigenvectors in their corresponding region.

APPENDIX E ASYMPTOTICS OF CANONICAL EIGENVECTORS AND NONLINEAR FOURIER COEFFICIENTS WHEN $|\lambda| \gg 1$

If $\lambda \in \mathbb{C}^+$ and $|\lambda| \gg 1$, then $\frac{1}{j(\omega-2\lambda)} = \frac{-1}{2j\lambda} + O(\lambda^{-2})$ and, taking the inverse Fourier transform, we can approximate $e^{2j\lambda t} u(t) = -\frac{1}{2j\lambda} \delta(t) + O(\lambda^{-2})$. Substituting into (51), (52), (53), for $|\lambda| \gg 1$ we obtain

$$V^2(t, \lambda) = \begin{pmatrix} 1 + \frac{1}{2j\lambda} \int_{-\infty}^{\infty} |q(t)|^2 dt \\ \frac{1}{2j\lambda} q^*(t) \end{pmatrix} + O(\lambda^{-2}). \quad (55)$$

A similar asymptotic expression can be derived for V^1 ($\lambda \gg 1$)

$$V^1(t, \lambda) = \begin{pmatrix} \frac{1}{2j\lambda} q(t) \\ 1 + \frac{1}{2j\lambda} \int_{-\infty}^{\infty} |q(t)|^2 dt \end{pmatrix} + O(\lambda^{-2}).$$

If $|\lambda| \rightarrow \infty$, q can be assumed zero in (21) compared to $j\lambda$. Thus $v(t, \lambda)$ approaches the boundary conditions at $t = \pm\infty$. Therefore

$$\begin{aligned} a(\lambda) &= \langle v^2(t, \lambda), v^1(t, \lambda) \rangle_s = \langle v^2(+\infty, \lambda), v^1(+\infty, \lambda) \rangle_s \\ &\rightarrow \langle v^2(-\infty, \lambda), v^1(+\infty, \lambda) \rangle_s = 1. \end{aligned} \quad (56)$$

Similarly, it is shown that $b(\lambda) \rightarrow 0$ as $|\lambda| \rightarrow \infty$.

APPENDIX F

SOLUTION OF THE RIEMANN-HILBERT PROBLEM

In Section VII, the inverse nonlinear Fourier transform was formulated as an instance of the Riemann-Hilbert factorization problem. Using hints given in the Appendix C, the resulting factorization problem can be solved in a simplified manner via an appropriate contour integration.

Dividing both sides of the projection equations (27)-(28) by $a(\lambda)(\lambda - \zeta)$, for parameter $\zeta \in C^+$, and integrating on the real axis $-\infty < \lambda < \infty$, we obtain

$$\begin{aligned} \frac{1}{2\pi j} \int_{\lambda=-\infty}^{\infty} \frac{V^2(t, \lambda)}{a(\lambda)(\lambda - \zeta)} d\lambda &= \frac{1}{2\pi j} \int_{\lambda=-\infty}^{\infty} \frac{\tilde{V}^1(t, \lambda)}{\lambda - \zeta} d\lambda \\ &+ \frac{1}{2\pi j} \int_{\lambda=-\infty}^{\infty} \frac{\hat{q}(\lambda) e^{2j\lambda t} V^1(t, \lambda)}{\lambda - \zeta} d\lambda, \end{aligned} \quad (57)$$

in which the integration is performed on the *open path* $z = \lambda + j\Im(\zeta)$, $-\infty < \lambda < \infty$. Without loss of generality, we can assume that ζ is real, as otherwise, we can shift the real line to the place of ζ , by replacing $f(z)$ with $f(z - j\Im(\zeta))$ and integrating from $-\infty + j\Im(\zeta)$ to $+\infty + j\Im(\zeta)$. The integration path is assumed to pass the singularity $\lambda = \zeta$ from above in all the integrals.

Cauchy integrals in (57) are computed from the residue theorem. The integration path $-\infty < \lambda < \infty$ can be closed in the upper or lower half-planes. To compute the first integral, we close the path in the upper half-plane and denote the resulting closed contour by $C_{-\zeta}^+$, i.e., the upper half plane and excluding the singularity $z = \zeta$

$$\begin{aligned} &\frac{1}{2\pi j} \int_{\lambda=-\infty}^{\infty} \frac{V^2(t, \lambda)}{a(\lambda)(\lambda - \zeta)} d\lambda \\ &= \frac{1}{2\pi j} \oint_{C_{-\zeta}^+} \frac{V^2(t, z)}{a(z)(z - \zeta)} dz - \lim_{R \rightarrow \infty} \frac{1}{2\pi j} \int_{Re^{j0}}^{Re^{j\pi}} \frac{\begin{pmatrix} 1 \\ 0 \end{pmatrix}}{a(z)(z - \zeta)} dz \\ &= \sum_{i=1}^N \frac{V^2(t, \lambda_i)}{a'(\lambda_i)(\lambda_i - \zeta)} - \frac{1}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \sum_{i=1}^N \frac{b(\lambda_i) e^{2j\lambda_i t} V^1(t, \lambda_i)}{a'(\lambda_i)(\lambda_i - \zeta)} - \frac{1}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \sum_{i=1}^N \frac{\tilde{q}(\lambda_i) e^{2j\lambda_i t} V^1(t, \lambda_i)}{\lambda_i - \zeta} - \frac{1}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \end{aligned} \quad (58)$$

where, in the second line, when $R \rightarrow \infty$, we have used the asymptotic values (55) and (56) in Appendix E. Note that we

assumed that eigenvalues λ_j are all *simple* zeros of $a(\lambda)$, i.e., no multiplicity.

To compute the second integral in (57), we close the integration path in the lower half-plane and denote the resulting closed contour by $C_{+\zeta}^-$, i.e., the lower half-plane and including the singularity $z = \zeta$

$$\begin{aligned} &\frac{1}{2\pi j} \int_{\lambda=-\infty}^{\infty} \frac{\tilde{V}^1(t, \lambda)}{\lambda - \zeta} d\lambda \\ &= \frac{1}{2\pi j} \oint_{C_{+\zeta}^-} \frac{\tilde{V}^1(t, z)}{z - \zeta} dz - \lim_{R \rightarrow \infty} \frac{1}{2\pi j} \int_{Re^{j2\pi}}^{Re^{j\pi}} \frac{\begin{pmatrix} 1 \\ 0 \end{pmatrix}}{z - \zeta} dz \\ &= -\tilde{V}^1(t, \zeta) + \frac{1}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \end{aligned} \quad (59)$$

The last integral in (57) is left uncomputed, because the boundedness of $e^{2j\lambda t}$ depends on the sign of t . For $t > 0$, we can consider $C_{-\zeta}^+$ which leads to the expression (58) multiplied by $u(t)$. For $t < 0$, we should inevitably consider $C_{+\zeta}^-$, where poles of $V^1(x, \lambda)$ are unknown. As a result, this integral is left untreated.

Using (58) and (59) in (57), we obtain an integral equation relating canonical eigenvectors V^1 and \tilde{V}^1 to $\hat{q}(\lambda)$ and $\tilde{q}(\lambda_j)$

$$\begin{aligned} \tilde{V}^1(t, \zeta) &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \sum_{i=1}^N \frac{\tilde{q}(\lambda_i) e^{2j\lambda_i t} V^1(t, \lambda_i)}{k - \lambda_i} \\ &+ \frac{1}{2\pi j} \int_{\lambda=-\infty}^{\infty} \frac{\hat{q}(\lambda) e^{2j\lambda t} V^1(t, \lambda)}{\lambda - \zeta} d\lambda. \end{aligned} \quad (60)$$

This is the first equation in the Riemann-Hilbert system (40). Since (60) holds for any $\zeta \in C$, evaluating (60) at $\zeta = \lambda_j$, $j = 1, \dots, N$, produces the third equation in (40). A similar integral equation for V^2 and \tilde{V}^2 gives the remaining equations.

REFERENCES

- [1] V. E. Zakharov and A. B. Shabat, "Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media," *Soviet Physics-JETP*, vol. 34, pp. 62–69, 1972.
- [2] G. P. Agrawal, *Nonlinear Fiber Optics*. Springer, 2006.
- [3] R. J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *IEEE J. Lightwave Techn.*, vol. 28, no. 4, pp. 662–701, February 2010.
- [4] G. Kramer, A. Ashikhmin, A. Wijngaarden, and X. Wei, "Spectral efficiency of coded phase-shift keying for fiber-optic communication," *IEEE J. Lightwave Techn.*, vol. 21, no. 10, p. 2438, 2003.
- [5] S. M. Moser, "Duality-based bounds on channel capacity," Ph.D. dissertation, ETH Zurich, Switzerland, 2004.
- [6] E. Ip and J. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *IEEE J. Lightwave Techn.*, vol. 26, no. 20, pp. 3416–3425, 2008.
- [7] K. Roberts, C. Li, L. Strawczynski, M. O'Sullivan, and I. Hardcastle, "Electronic precompensation of optical nonlinearity," *IEEE Photonics Techn. Lett.*, vol. 18, no. 2, pp. 403–405, 2006.
- [8] Y. Cai, J. Cai, A. Pilipetskii, G. Mohs, and N. Bergano, "Spectral efficiency limits of pre-filtered modulation formats," *Opt. Exp.*, vol. 18, no. 19, pp. 20 273–20 281, 2010.
- [9] A. Tarighat, R. Hsu, A. Shah, A. Sayed, and B. Jalali, "Fundamentals and challenges of optical multiple-input multiple-output multimode fiber links," *Communications Magazine, IEEE*, vol. 45, no. 5, pp. 57–63, 2007.

- [10] L. Beygi, E. Agrell, M. Karlsson, and P. Johannisson, "Signal statistics in fiber-optical channels with polarization multiplexing and self-phase modulation," *IEEE J. of Lightwave Techn.*, vol. 29, pp. 2379–2386, 2011.
- [11] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform*, 1st ed. Society for Industrial and Applied Mathematics (SIAM), 2000.
- [12] L. D. Faddeev and L. A. Takhtajan, *Hamiltonian methods in the theory of solitons*. Springer Verlag, 2007.
- [13] T. Tao and C. Thiele, "Nonlinear Fourier analysis," UCLA Department of Mathematics, <http://www.math.ucla.edu/~thiele/lecturenotes/nlft.pdf>, Tech. Rep., 2006.
- [14] E. Fermi, J. Pasta, and S. Ulam, "Studies of nonlinear problems," *Los Alamos document LA 1940*, 1955.
- [15] N. J. Zabusky and M. D. Kruskal, "Interaction of "solitons" in a collisionless plasma and the recurrence of initial states," *Phys. Rev. Lett.*, vol. 15, no. 6, pp. 240–243, 1965.
- [16] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, "Method for solving the Korteweg-de Vries equation," *Phys. Rev. Lett.*, vol. 19, no. 19, pp. 1095–1097, 1967.
- [17] P. D. Lax, "Integrals of nonlinear equations of evolution and solitary waves," *Comm. Pure Appl. Math.*, vol. 21, no. 5, pp. 467–490, 1968.
- [18] V. E. Zakharov and F. Calogero, *What is integrability?* Springer-Verlag, 1991.
- [19] A. C. Singer, "Signal processing and communication with solitons," Ph.D. dissertation, Massachusetts Inst. of Technology, 1996.
- [20] M. Reed and B. Simon, *Methods of Modern Mathematical Physics: Vol I, Functional Analysis*. Academic Press, 1980.
- [21] L. C. Evans, *Partial Differential Equations*. American Mathematical Society, 2010.
- [22] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, "The inverse scattering transform - Fourier analysis for nonlinear problems," *Studies in Applied Mathematics*, vol. 53, pp. 249–315, 1974.
- [23] A. S. Fokas, "A unified transform method for solving linear and certain nonlinear PDEs," *Proc. Royal Society London A*, vol. 453, pp. 1411–1443, 1997.
- [24] E. M. Stein and R. Shakarchi, *Complex analysis*. Princeton University Press, 2003.
- [25] M. Klaus and J. Shaw, "On the eigenvalues of Zakharov-Shabat systems," *SIAM journal on mathematical analysis*, vol. 34, no. 4, pp. 759–773, 2003.
- [26] N. Muskhelishvili and J. Radok, *Singular integral equations*, 2nd ed. Dover Publications, 2008.
- [27] T. Tao, *Nonlinear dispersive equations: local and global analysis*. American Mathematical Society, 2006.